Bcep176 AND Bglu421 – TWO NOVEL PHAGES CONTRIBUTING TO THE UNDERSTANDING OF PATHOGENICITY AND DIVERSITY IN *Burkholderiacae*

A Senior Scholars Thesis

by

LINET MERA

Submitted to the Office of Undergraduate Research
Texas A&M University
In partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

April 2006

Major: Genetics & Biochemistry

Bcep176 AND Bglu421 – TWO NOVEL PHAGES CONTRIBUTING TO THE UNDERSTANDING OF PATHOGENICITY AND DIVERSITY IN *Burkholderiacae*

A Senior Scholars Thesis

by

LINET MERA

Submitted to the Office of Undergraduate Research
Texas A&M University
In partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor: Elizabeth Summer
Associate Dean for Undergraduate Research: Robert C. Webb

April 2006

Major: Genetics & Biochemistry

ABSTRACT

Bcep176 and Bglu421 – two novel phages contributing to the understanding of pathogenicity and diversity in *Burkholderiacae* (April 2006)

Linet Mera
Department of Biochemistry and Biophysics
Texas A&M University

Research Advisor: Dr. Elizabeth J. Summer
Department of Biochemistry and Biophysics

Phage, although but a fraction of the size of bacteria, can, by lysogenic conversion, transform a harmless bacterium into a ruthless pathogen (3). This paper will discuss two new lysogenic dsDNA tailed phages of *Burkholderia*. Bcep176 was induced from *Burkholderia multivorans* ATCC17616, thought to be found in pulmonary infections (9), and Bglu421, from a rice seed isolate of *B. glumae* BG 10421, known to be the causative agent for rice seed rot (19). The capsid-to-tail proteins and tail assembly cassette of Bcep176 demonstrated significant modular mosaicism to homologs in lambdoid phages phiE125 and phi1026b – phages involved in the study of bacterial bioweapons (13, 26). Bglu421 exhibited modular mosaicism in tail assembly cassette and putative morons to homolog counterparts in a previously unannotated prophage of *Burkholderia pseudomallei* Pasteur - causative agent of melioidosis in a variety of animals (12). Further analysis of both phages indicates siphophage morphology and lambda-like proteins in Bcep176 and a putative pathogenicity factor in Bglu421. The importance of genomic sequencing and analysis of bacteriophages lies in their role of influencing the pathogenicity of their host (5), and their potential for use as cures of bacterial disease (1).

ACKNOWLEDGEMENTS

I would like to acknowledge my research supervisor Dr. Summer for all her guidance and support throughout the course of this research and thesis-writing process. Without her infinite patience and wisdom none of this would have been possible.

I would also like to thank Dr. Gonzalez and his lab at Texas A&M for their work in isolating *Burkholderia glumae* BG10421 and the phages we have used. I am also deeply indebted to LiLi Niu for facilitating management of the tables upon tables of data herein presented.

Thanks are also in order for all my friends and colleagues in lending their support through all the trials and tribulations of research and college. In particular, I'd like to thank my following closest friends: Ula for always being there for me even if I was always at the lab on nights and weekends; Andrea, Sabrina, Erika, Stephanie, Niki, Stephen, Lucki, and Julia for keeping my head up when I was most worried; Ian, Brian and John for being my companions in this insane quest for international networking and cross-disciplinary research that impacts the world.

To Dr. Hu, Dr. Wild, and Dr. Young I express my most sincere appreciation for their guidance and encouragement to continue pursuing a career in the sciences.

I'd like to extend my gratitude to my parents, Luis and Luz, for always encouraging, guiding and supporting me through my failures and successes in college and in life. My brother also deserves mention, his zeal for knowledge reminds me of myself and keeps me on course.

Lastly, I'd like to thank the National Science Foundation for the monetary resource without which this research would not be possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION[1]: BACTERIOPHAGE - A FORCE TO BE RECKONED WITH

Bacteriophage background. Phages are properly known as bacteriophages, bacterio- referring to bacteria and -phage from the Greek word 'phagos' meaning to eat. A bacteriophage is a type of virus that can attack eubacteria. Phages are extremely small, those discussed in this text are only about 75 x 300 nm while a common E. coli bacterium is about 2x1μm. It is estimated that for each bacteria that exists in our biosphere there are 10-100 phages, totaling up to 10^{31} phages on earth (3). Bacteriophages have a distinct structural appearance: an icosahedral protein head (capsid) that protects the genetic material within; a hollow tail through which the genetic material can pass; a portal that connects the tail and capsid; tail fibers key to identification of and absorption to the host cell; and a baseplate that connects the tail fibers to the tail and contains a structure (tail spike) through which to inject the genetic material into the host cell (see Fig. 1)(6).

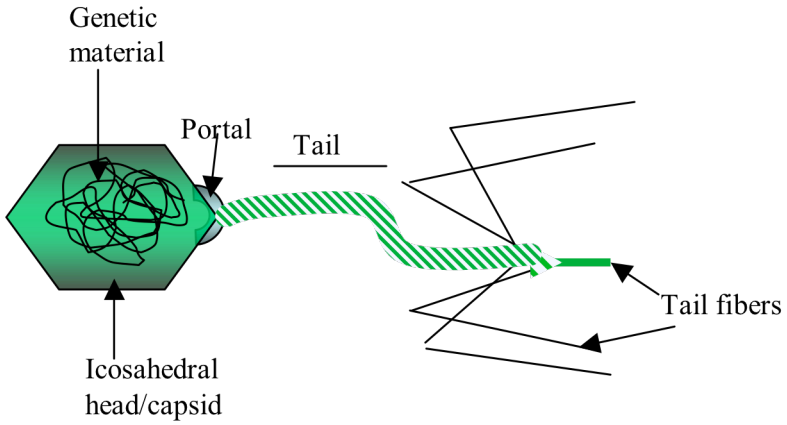

FIG. 1. Basic siphophage structural components

[1]This thesis follows the style of Journal of Bacteriology.

Phages differ from one another in the method in which their genetic material is encoded: double-stranded (ds) or single-stranded (ss) DNA or RNA There are three general morphotypical categories of dsDNA tailed (caudovirii) phages: siphophage, podophage, myophage. The difference between the morphologies lies mainly in their tail (flexible, short, contractile – respectively). Caudovirii can also be characterized in accordance with their shared similarities to model phages, such as the well-studied enteric phage lambda. In particular, lambda-like phages fall under the siphophage (flexible tail) category and have very discrete compartmentalization of their genome (Fig. 2). The dsDNA group of tailed phages can be further classified according to life cycle: lysogenic (temperate) or lytic (virulent). Both types of phages can identify and infect the host in the same manner: detection and adsorption of the host via tail fibers and subsequent injection of DNA. Upon injection, a lytic phage would immediately begin replication, transcription and translation of its genes, progeny assembly, and lysis(death) of the host cell via a perfectly timed lysozyme mechanism at the host's wall. After injection, a lysogenic phage would be able to choose between lying dormant in the host cell, as a plasmid or integrated into the genome (prophage), and pursuing a lytic phage life cycle. Once the phage integrates into the host DNA or is circularized while being replicated in concert with the host DNA, severe environmental stress would be required for the phage to excise itself from the host (lysogen) genome in order to continue the lytic cycle. The general sequence of events in a lysogenic life cycle is represented in Fig. 3. (6)

Both phages discussed herein are inducible, dsDNA lysogenic siphophage. Once integrated into the host genome (lysogenized) or within the host as a plasmid, a prophage can suffer the same fate as its neighboring DNA: spontaneous mutation and elimination from the host as extraneous DNA. A prophage that is able to excise itself from the host DNA and begin the lytic cycle is considered fully functional. However a phage that has undergone enough mutational decay so that it cannot excise itself from the host is considered to be a nonfunctional prophage. In order to prevent annihilation and deactivation, a phage has several self-preservation options: it can express DNA-

modification elements in order to make its foreign DNA ‘impervious’ to detection, prevent multiple-phage infection, contain fitness cassettes that can benefit the host or produce toxins harmful to the host if ever excised. (3)

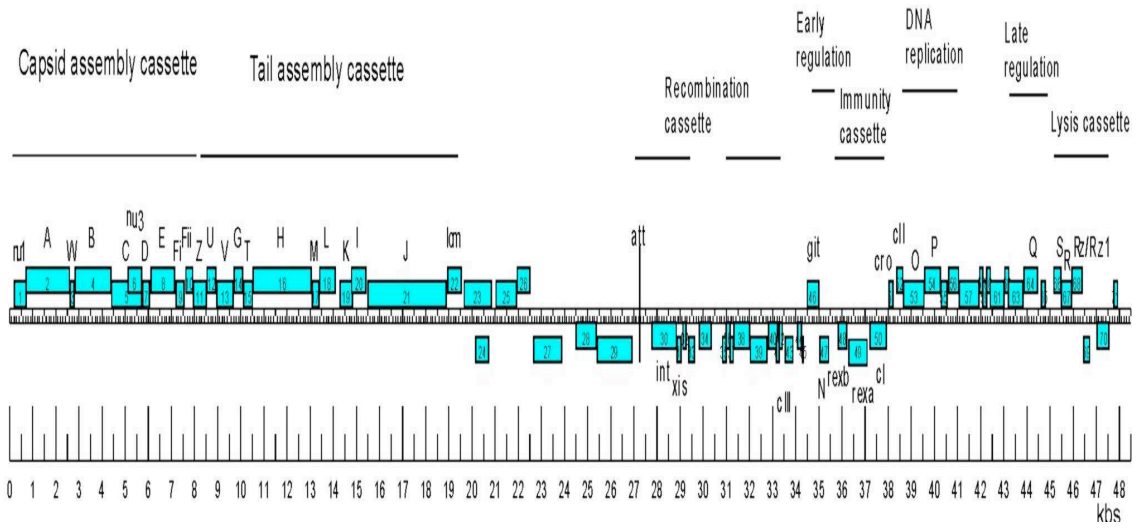

FIG. 2. Lambda genome (from NC_001416 2006)

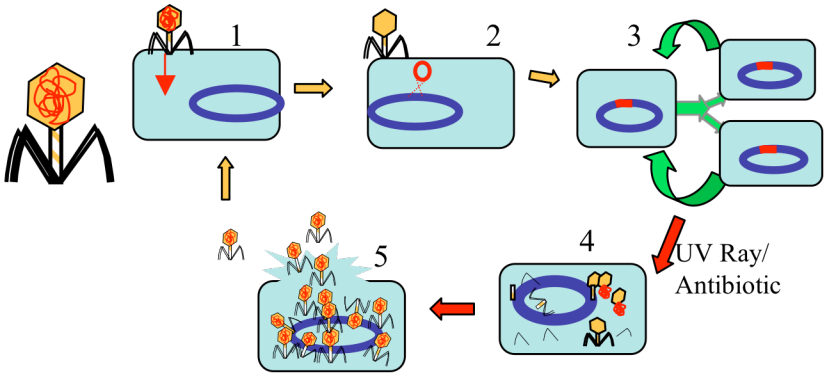

FIG. 3. Lysogenic phage life cycle

1)Phage finds host using its tail fibers, adsorbs to host, and injects DNA
2)Phage DNA circularizes and then integrates into host genomic DNA
3)Prophage is replicated along with host genomic DNA, this step can be repeated interminably until host cell lysis is induced.
4)Once host cell is damaged, prophage DNA excises itself from host genome and begins replication and progeny formation.
5)The host cell is lysed and the cell dies due to loss of potential difference between its outer and inner membrane. Hundreds of fully functional phage are released and are free to infect another bacterial cell.

Phage evolution, bacteria and pathogenicity. Despite several methods of general classification, phage diversity is still too broad to be discretely categorized. Unlike eukaryotes and prokaryotes, excepting vertical gene transfer, phage do not seem to evolve by small changes in their genome over measurable, predictable periods of time. Instead, it is proposed in the modular theory (2, 18, 22, 25) that phage evolve via exchange (horizontal transfer) and duplication (vertical transfer) of discrete sets of alleles called modules – a theory which continues to be well-supported with genomic sequencing (3). When a phage shares several modules with another phage, in noticeable contiguous sets scattered over the genome, the phages are said to show mosaicism to one another. In addition to modular exchange, it has been observed that some phage carry around 'extra' genes that contain their own promoter and terminator sequence, but are non-essential to phage life cycle (3). These lysogenic conversion genes, or 'morons' (17) seem to play a role in making the phage more fit for survival within its host by influencing lysogen fitness. Morons can encode for virulence (pathogenicity) factors that can allow the lysogen to perform many functions, including resistance to host defenses and increased or acquired ability to harm and/or destroy the host (3). Since a phage's life cycle is inextricable from that of its bacterial host, phage genomics and evolution are also closely tied to bacterial genomics and evolution (3, 6).

Other than lysogenic conversion, there are four other ways in which lysogenic phage can affect host fitness - genome rearrangement, gene disruption, and protection from lytic infection (3). For the purpose of this research, focus will be centered on lysogenic conversion. Lysogenic conversions leading to disastrous human pathogens are numerous and include the following more infamous examples (listed in Table 1): conversion of common *Escherichia coli* to *E. coli* O157:H7 (20), non cholera-causing *Vibrio* bacteria to *V. cholera* (24), and harmless *Corynybacterium* to *C. diphtheriae* (16).

Thus sequencing and bioinformatic analysis of new lysogenic phage allows us to elucidate their contribution to their host's fitness and furthermore its pathogenicity, and add to the knowledge of phage and bacterial evolutionary and pathological diversity (3, 6).

TABLE 1. Infamous lysogenic conversions

Lysogen	Pathogenic function provided by prophage	Prophage
E. coli O157:H7	Shiga toxins	Various depending on strain
Vibrio cholera O1 &O139	Toxin-corregulated pilus (TCP) & Cholera toxin (CT)	Ctxf
Corynebacterium diphtheriae	Diphtheria toxin (DT)	Corynephage beta

***Burkholderiacae* – burgeoning with diversity.** Both novel phages sequenced are inducible lysogenic dsDNA tailed phage newly isolated from two different *Burkholderia* bacteria.

Burkholderia multivorans and *glumae* bacteria are both part of a group of *Burkholderia* bacteria known as the *Burkholderia cepacia* complex or BCC (10) that bear importance to the medical, agricultural, and environmental world. BCC are notorious for their involvement in causing cystic fibrosis infections (11), rice blight (19), melioidosis and glanders (14), their use as PCB degraders (15), and misuse as potential biological weapons of terror (7, 8). A better understanding of *Burkholderia* phage may contribute to the understanding of *Burkholderiacae* diversity as a whole and bring us closer to revealing the evolutionary connections between pathogenic and non-pathogenic strains. To this end, two novel lysogenic phages of *Burkholderia* were sequenced, Bcep176 and Bglu421. I am reporting on genomic analysis of these phages, particularly on their relationship to previously described lysogenic prophages of *Burkholderia* (13, 26) as well as one previously unannotated prophage.

METHODS

Media and bacterial strains. NBY broth is 0.8 % nutrient broth, 0.2 % yeast extract, 0.2 % dibasic; K_2HPO_4, 0.05 % monobasic $K_2H_2PO_4$; NBY agar is NBY broth with 0.02 % unpurified agar; TNB is 0.5 % tryptone, 0.25 % yeast extract, 0.1 % dextrose, 0.85 % NaCl, 0.4 % KNO_3.

Bacterial strains used in this report are *Burkholderia multivorans* ATCC17616 (Stanier, soil USA) and *Burkholderia glumae* BG10421 (Gonzalez, TX rice seed). Frozen stocks were made from log phase cells for cryo-preservation in 40 % TNB and 60 % sterile gylcerol.

Phage isolation. 15 mls of NBY broth was inoculated to an OD_{550} of 2 from a fresh overnight streak plate (grown at 24^{o}C on NBY agar). This was used to inoculate 50 mls of NBY broth (in a 250 ml Erlenmeyer flask) to an OD_{550} of 0.1 (approximately a 1:10 dilution). This was incubated in a shaking water bath (225 rpm) at 30^{o}C. When the culture reached an OD_{550} of 0.4, lysogens were induced by the addition of mitomycin C (2 μg/ml), resuspended in 10 mM Tris pH 8. OD_{550} measurements were taken every hour until lysis. At this time, chloroform was added to a final concentration of 0.1 % to complete the lysis process. The lysate was then purified by pelleting bacterial debris (5k rpm JA-20 rotor) and filtering the supernatant through 0.2 μm Steriflip (Millipore) tube. Liquid lysates were stored at 4^{o}C. (see Fig. 4 for diagram)

Library preparation. Phage DNA was purified from 10 mls of lysate utilizing the Promega Lambda Wizard Purification kit following the manufacturer's specifications. The DNA was hydrosheared into 2 kb average size fragments using GeneMachine Hydroshear device and size-selected by gel extraction following QIAEX II Agarose Gel Extraction (Qiagen). Ends of the DNA were then repaired with DNA Terminator End-Repair kit (Lucigen) and ligated into pSmart HC Kan vector (Lucigen). The ligation reaction was electroporated into *E. cloni* (Lucigen) cells according to manufacturer's specifications.

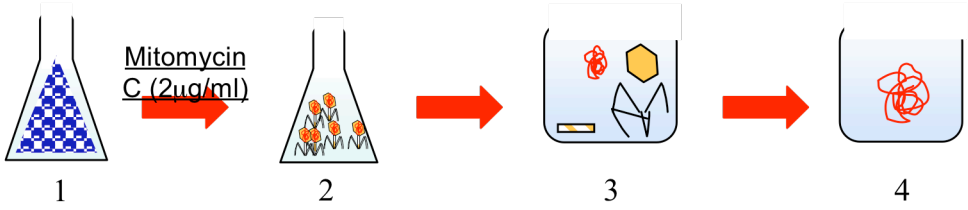

FIG. 4. Isolating phage DNA

1) *Burkholderia* strains initiallyisolated from soil samples are purified and grown to OD=0.35 2) Prophage are induced using mitomycin C an antibiotic. 3) DNAase/RNAase added to remove bacterial nucleic acids. Phage particles are then disrupted with guanidinium HCl. 4) Purified phage DNA!

Plasmid isolation. Transformants were selected by plating on LB kanamycin (30 mg/L) agar plates and arrayed into 96-place deep well blocks containing 1.46 ml LB kan broth for plasmid preparation. The blocks were covered with microporous sealing film (USA Scientific) and incubated at 37°C overnight with aeration at 225 rpm. The bacteria were pelleted by centrifugation at 1k rpm for 10 mins. The supernatant was then poured off, and the bacterial pellets were stored at -20°C until plasmid isolation. Plasmids were then isolated from the bacterial pellet using the Qiagen 96 R.E.A.L. prep procedure with the BioRobot 3000 (Qiagen). These DNA plasmid pellets were resuspended in 150 µl of 5 mM Tris pH 8.

Sequencing reactions. Reactions took place in a 96-well format. For each reaction, 1 µl of plasmid DNA was used as the template in a 7.5 µl sequencing reaction with left and right primers using Big Dye 3.1 (ABI). Reactions were incubated using thermocycler conditions: 99 cycles of 96°C– 60°C–60 °C for 10"–5"– 4' (respectively). Sequencing reactions were purified by isopropanol precipitation. Reactions were subject to capillary sequence analysis - either at Seqwright (Houston, TX) or at Dr. Engyu No's Laboratory for Plant Genome Technology in Texas A&M University's Institute for Plant Genomics and Biotechnology.

Sequence editing. Resulting sequences were end-trimmed and vector sequence was removed using Sequencher (Gene Codes). Trimmed sequences were then

hierarchically assembled yielding contiguous segments of sequences (contigs). The contigs were then edited for ambiguities by comparing the consensus sequence against chromatograms of raw data and against each individual sequence contained in the contig. Gaps and low coverage areas were filled for Bcep176 by primer walking. Bglu421 sequence is at the initial shotgun stage.

Gene prediction. Protein-coding genes were initially predicted with Genemark Heuristic Model (http://opal.biology.gatech.edu/GeneMark/heuristic_hmm2.cgi). Comparisons between Genemark predicted proteins and database sequences were performed with BlastP (October 24, 2005 for Bcep176 and March 24, 2006 for phiPasteur and Bglu421) (http://www.ncbi.nlm.nih.gov/blast/) using 0.1 as the lowest expect value (homologies with 1.00E-03 or more are considered insignificant hits).

Genomic sequence annotation. The program Artemis v.7 was utilized for visualization and manual editing/addition of open reading frames. Once corrected, all contiguous sequences along with their predicted genes were processed through a Blast program in order to confirm and update protein homologies and find new ones (among created genes). In order to facilitate the annotation process a Blast Parser or organizer was utilized (http://dimer.tamu.edu/young/genomics/annotation/). Annotations were assigned based on the following criteria: proteins were assigned a functional annotation if they exhibited significant similarities(usually equivalent to or less than E-4) to proteins of known functions or had distinctive protein-gene organization characteristics; no homologies in Genbank - hypothetical novel protein; homologies to only hypothetical proteins – hypothetical conserved; homologies to hypothetical phage proteins – hypothetical conserved phage protein. After several iterations of the above-stated procedure the annotated sequence of Bcep176 was submitted to Genbank (DQ203855) Bglu421 will also be submitted upon completion. (See Fig. 5 for diagram of whole shotgun genome approach)

Identification of related prophages. Blast output data was organized in order to detect homolog clustering from a single source. If these were from a bacterial genomic sequence, the cognate sequences were extracted from the database submission.

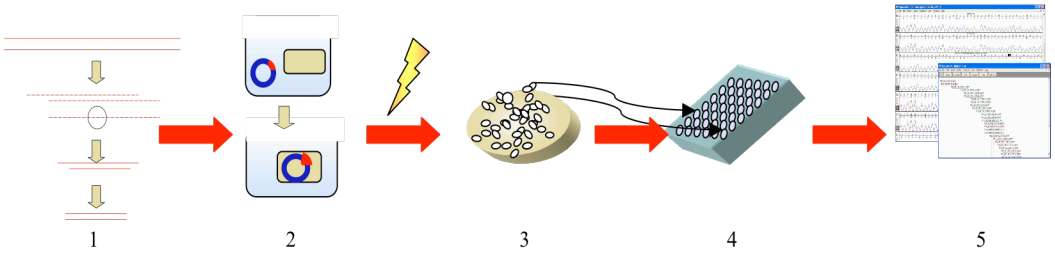

FIG. 5. Whole shotgun genome sequencing diagram

1)DNA is sheared into 2kb fragments and end-repaired. 2) The hydrosheared(HS), end-repaired DNA is then ligated into a pSmartHCKan plasmid vector which is then transformed by electroporation into *E.cloni* (Lucigen) cells. 3) The E.coli are then plated on LBKan, colonies that appear contain the plasmid (and a Kanamycin resistance cassette). 4)A 96-well block containing LBKan was then inoculated with one colony/well. The cultures were allowed to grow overnight. The bacteria was then pelleted and the plasmid isolated from the bacterial debris (alkaline lysis prep). 5)The plasmids are then used as templates for sequencing with left and right primers. The sequencing results are assembled and analyzed using several automated analysis programs and many hours of scrutinizing the computerized output for incongruities.

RESULTS

Phage isolation. Bcep176 and Bglu421 are both newly isolated phages. Bglu421 was obtained from *Burkholderia glumae* BG10421, while Bcep176 was obtained from *Burkholderia multivorans* ATCC17616. Bglu421 DNA was isolated and purified as described, Bcep176 was also isolated in the same manner, although the insert library was readily available to proceed to inoculation into 96-well block format. At this point, Bcep176 is already available on Genbank (DQ203855) with 14x coverage and Bglu421 is in initial assembly stages.

Physiological parameters of Bglu421. The doubling time of *B. glumae* was determined to be about 2.5 hours at 30^{o}C. Under these same conditions lysis was estimated to occur was estimated to be between 3 and 11 hours after addition of inducing agent. Another unrelated phage Bglu191, was isolated from another *B. glumae* strain grown in similar conditions, had an estimated lysis time of 4 hours.

Morphology. Transmission electron microscopy (TEM) images (Fig. 6) reveal that Bcep176 has a well-defined flexible tail indicative of siphophage morphology; the results were supported by bioinformatic analysis of its structural cassette. The icosahedral head diameter is ~75 nm and tail ~300 nm. A TEM of Bgl421 is not available at this time.

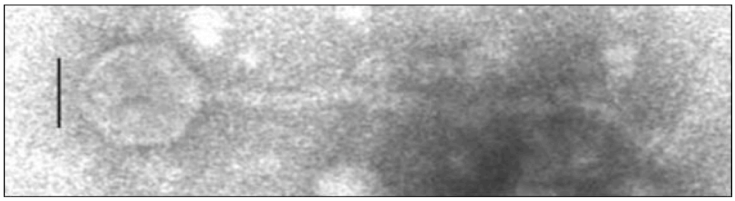

FIG. 6. Bcep176 TEM.
Negatively stained transmission electron micrograph of Bcep176.
Bar, 50nm.

Bcep176 analysis. *Genome structure.* Bcep176 was determined to have 44,856 bps of unique coding sequence. The genomic termini were found to be 3'extended COS ends. This was determined by PCR using ligated phage DNA as a template to PCR primers E (ACGTGGTTGACTGGTTGTTG) and B (ATTGGATCCCTTCTGCACAC). Sequencing of this PCR product produced 10 nts of sequence (GGAACGTGCG) not present in the shotgun assembly. This data indicates that the genomic ends consist of 10 nts of 3' extended COS ends. This data was supported following gene annotation of the Bcep176 TerL homolog [79] which clusters strongly with TerL homologs from other phages possessing 3' extended COS ends (4).

Genome overview. Bcep176 is predicted to encode 81 gene products (gp), 28 of which have no database homologs. A summary of predicted proteins can be found in Table 2.

Integration cassette. Integration modules are involved in the process of inserting the phage DNA into the host genome as a prophage. One of the proteins involved is the endonuclease protein [01], designated as such by its high expect value of 2.00E-45 to *Burkholderia thailandensis* E264 gi(gene ID from Genbank)|83717854| 59R and clustering to other endonucleases.

Capsid assembly cassette. Capsid assembly involves a variety of proteins, those that are actually included in the final structure, those that chaperoned capsid monomers, and those like scaffolding proteins that are integral structures to creating a capsid but are nowhere to be seen in the final product. Bcep176 contains several of these: portal protein [78], determined by its strong relation to *Burkholderia thailandensis* E264 gi|83716795| head portal protein with the highest expect value possible of 0; Clp-related prohead protease [77], defined as such by a similarity strength of 4.00E-114 to *Burkholderia thailandensis* E264 gi|83718056| ClpP protease; major capsid protein [76] was annotated by its strong expect value of 0 to *Burkholderia thailandensis* gi|83716436| phage major capsid protein of the HK97 family; head-tail adaptor protein [73], annotated

TABLE 2. Summary of phage Bcep176 coding regions and predicted proteins

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp01	-	123	512	389	[*Burkholderia thailandensis* E264] gi|83717854|ref|YP_439239.1| 59R, 112, , 2.00E-45: 81% over 85aa***	restriction endonuclease	-	MSKRLTTLKPRVQSLAASRVATMTPGSWR AGKSSSAARGYGYEWQKLRAAHLAKHPH CVFCLRDLGMAGWSPAEVVLACAARGIG EPLGTIGDHIVAHRGDRRLQLDPTNIQTLC KTHHDSEKQRLERG*
Bcep176gp02	-	730	1320	590	[*Burkholderia vietnamiensis* gi|67547316|ref|ZP_00425220.1| phage-related hypothetical protein,197, 2.00E-65: 70% over 135 aa***	hypothetical conserved phage protein	-	MSEIEFKSAFDAVRFALAYSTQQYGETMM AKRLRGESLGSGMGLVGVDGAGQAGMIR AKVWELPELHRAVIIARAAPRDLPCSCGKS CCSGSQPNPEWDAAISWLTDASAAYCSGF SHYRVRRAIIEGVFGVKRSLADVAEDCDA HVNTVSKQNAAVRKWLEGNSKTGAVGVI DAAWAAIERKLSTAGLLKNCEIA*
Bcep176gp03	-	1366	1614	248	[Bacteriophage phiE125] gi|17975227|ref|NP_536422.1| hypothetical protein phiE125p66, 86, 3.00E-05: 36% over 26 aa*** [[Bacteriophage phi1026b] gi|38505458|gb|AAR23227.1| gp76, 86, 8.00E-05: 36% over 26 aa***	phiE125 gp65-like, phi1026bgp 76-like	-	MSAAACVLYGDVPEPLLISAIRHRDSVTGA ELIAFDECPFSGEITETEHGVQIAFPWPRNR TLRHAIGDWLTHYGINFTVVM*
Bcep176gp04	-	1611	1949	338	[Bacteriophage phi1026b] gi|38505457|gb|AAR23226.1| gp75, 114, 6.00E-21: 55% over 54 aa***[Bacteriophage phiE125] gi|17975226|ref|NP_536421.1| hypothetical protein phiE125p65, 114, 6.00E-21: 55% over 54 aa***	phiE125 gp64-like;phi102 6bgp75-like	-	MILTVQLPAGRHSFKRKHGMGPAISSEMH RPLVTTVYRIARIPTVKRQLLAVVEVDAFIP ERHRTHIAPSDPRWVRPGVLRTKAYWIDN KKSRALGQFLASDALEVHLEDEA*
Bcep176gp05	-	1946	2389	443	[Bacteriophage phi1026b] gi|38505456|gb|AAR23225.1| gp74, 141, 1.00E-60: 74% over 109 aa***[Bacteriophage phiE125] gi|17975225|ref|NP_536420.1| hypothetical protein phiE125p64, 112, 2.00E-57: 87% over 97 aa***[*Burkholderia thailandensis* E264] gi|83716439|ref|YP_439233.1| gp74, 151, 4.00E-39: 56% over 83 aa***	phiE125 gp63-like;phi102 6bgp74-like	-	MKRSGFKPRKKPMARGSWSRKSSPLPEQA PRKKLMARRAKRPTVAEGSKYLAACRGE PCYLRVPGVCRLNPIDDTVVPCHSNQGRH GKAGAMKAKNEFTVPGCAACHAWLDQN RAGTPKQVKFDVWNRAFEEWAPVRARK MGEANCQ*
Bcep176gp06	-	2386	2967	581	[*Burkholderia xenovorans* LB400] gi|91782754|ref|YP_557960.1| hypothetical protein Bxe_A3071, 200, 1.00E-13: 30% over 56 aa***	hypothetical conserved protein	-	MSKRPWQKWSEQEEADLRRVWTADGPIK QYLDLFNGRSMHAVLAHGADMGLGERPN RNARAGHPTESAILRALKLGPMDSLELSA KTGISRRTVMKHLKALHAKRHVYVSAWE RFGASGYPARIYALGNRKDAPRPPARTPGQ KWQDRMSHLKKHRPDEYMRVMARRRAN AQRRTGTAKRDIAAQALFGRATA*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp07	-	2964	3434	470	[*Burkholderia pseudomallei* Pasteur] gi|67754135|ref|ZP_00493050.1| hypothetical protein BpseP_02003522, 126, 4.00E-25: 53% over 68 aa***	hypothetical conserved protein	-	MAGNKKPRKAYRPRAVRRTAGFDVLERRTPMDGDQKTDLGI AYYMALNEMTNGRGTEEHWSTVACALNIALVIAETGPGLDS ISIIKSALAGAVRARDRAARVGKWGFDGDALIDIRIALEIHDA QMATVSKAAILKALGEVHRRIDAGEVFKEAA*
Bcep176gp08	-	3434	3997	563	[Phage phiV10] gi|89152461|ref|YP_512300.1| hypothetical protein PhiV10p45, 114, 2.00E-38: 68% over 77 aa***	hypothetical conserved phage protein	-	MTKRTAWPMRVDAGTKNVGTARVHDDSRPKMTTAQRAIFE ATGNRPHVDAAFDDISDGIDAPPVLTPAYRQPDAKTRMQAL GRLKAGAMNKTEQRYADHLEARKQAGEIVWYRFEGIKFRL ADNTFYTPDFTVMLANGQLEAHEVKGHWQDDARVKVKVA ADQYPVRFVAVKAAAVRDGGGWQTEEF*
Bcep176gp09	-	3994	4170	176	***no significant hits***	hypothetical novel protein	-	MTVDECRERFMAAVRDARAGRNGKARELIAAVRERFGDAA AETARRELRNYVDSDKKA*
Bcep176gp10	-	4167	4814	647	[*Burkholderia pseudomallei* Pateur] gi|67756056|ref|ZP_00494944.1| hypothetical protein, 203, BpseP_02001724, 6.00E-40: 41% over 80 aa***	hypothetical conserved protein	-	MTPSDYEEFSNLMAGVFAFYKRDVSEFALGVWWAAMKPYD LAAVTDALGRHSVNPDSGQFMPMPADIVKMLGGSTQDAALV AWAKVDRAVRSCGTYNSVVFDDALIHRVIVEMGGWVLIGG KSEDDWPFVRNEFVNRYRGYKMRSETPEYLPVLIGMAEAQ NNRTGHKSQPPVLIGDAHAAHRVMLGGQDKPMLGFVRMAP ELAANRPLPQLGAA*
Bcep176gp11	-	4811	5617	806	[*Aeromonas hydrophila*] gi|44151584|gb|AAS46719.1| replication protein, 268, 1.00E-17: 27% over 76 aa***[*Escherichia coli* O157:H7] gi|15831022|ref|NP_309795.1| hypothetical protein ECs1768, 346, 2.00E-17: 29% over 93 aa***	phiE125 gp60-like; phi1026bgp71-like	-	MSLDATTWARHQKVGKGPAKAVLMALADYANEDFITYPSV ETLVAWTEQDRKTVLANLDRLKEAGWITDTGERAGRTRQV VVYQINVDRGVEVKIGPRELLTGPKSEPSQNRNSSENGTVPN STGNSPNFDGKQSQKPPETVPNLGHRTVGTVENSRNSVGAR GTRLPDDWVLTKALGEWALAEQPTWTVEHVRKVAEKFADH WRAQPGQKGRKTDWAATWRNWVRTEKPLAGAAVGGGNK QESLEARNHEIARRAAEKFKAQNGGAA*
Bcep176gp12	-	5614	5712	98	***no significant hits***	hypothetical novel protein	-	MAMNAIRPYAEAGELNVSSIGPYALTLIVEAV*
Bcep176gp13	-	5723	5899	176	***no significant hits***	hypothetical novel protein	-	MDQHQASTDLEQLAALARQPTKLPGGLAEWLARRPTYPRIC LSCGAVESLDGSVPCDH*
Bcep176gp14	-	5899	6345	446	[Bacteriophage phi1026b] gi|38505447|gb|AAR23216.1| gp65, 8.00E-50: 63% over 92 aa***[Bacteriophage phiE125] gi|17975217|ref|NP_536412.1| hypothetical protein phiE125p56, 146, 1.00E-52: 67% over 99 aa***	phiE125 gp55-like; phi1026bgp65-like	-	MNIIDAAYAVVHDYPGGSESLAPRLGMSGAMLRNKVNPNN DTHKLSLAEAARIADVTNDDRILEAWARERGYALVEIPSPEN CSDGEIIELMAKTWETNGDIGKEIIRTFEDGRVEKHEVLRVKE RTWKHFQMLLGLVSRIEGMAEEQ*
Bcep176gp15	+	6534	6770	236	***no significant hits***	hypothetical novel protein	-	MEEAKVTFNFIGATASSIPTNIPDALKFRAGFTPNDGDVISFEG LPGAFLIHSREFQIRSDGVVDVVLNVDVAQPAKG*
Bcep176gp16	+	6745	7026	281	***no significant hits***	hypothetical novel protein	-	MWRSQQKANVDYRFDAPRRSKRLYEGRSSPRVIANAPSFEIT QWRPETETAHFIRHVESPCRSRSPRATSSAGRISHRTRRAYRR FMRPPFVQ*
Bcep176gp17	+	7038	7379	341	[*Azoarcus* sp. EbN1] gi|56315608|emb|CAI10253.1| hypothetical protein, 106, 7.00E-08: 34% over 34 aa***	hypothetical conserved protein	-	MNEKSNIGNIASGRRIAQCWYSATEYDQIKSVMDDGHKLPS RYEDWRAGAEQREDQVRRAGGTPVRIAFDLAEFRRFCAHFR VPLDADTRSKFAALKSQIDAEAGSEPGTGVH*
Bcep176gp18	+	7463	7834	371	[*Synechocystis* sp. PCC 6803] gi|38423900|dbj|BAD02108.1| slr6051, 126, 2.00E-08: 29% over 35 aa***	hypothetical conserved protein	-	MKRDMDIVRRIALAAEDLQYGYHLTGLDDVAPEVFGIHVIW MKEAGLVHAHVSEYLSPLDDPPDASVIRLTWSGCEFVDAAR SDTIWNKAKTTLIKPAASFSFQILREWLAAEIKQGLPTLRG*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp19	+	7842	8315	473	***no significant hits***	hypothetical novel protein	-	MSCKSHATDANTRFSAGNSARSAALTANRASTTQPARSNQL EHKISRPRSNASMSPDSSALIDLAPIPATEPRVICSSVMSRHLF LLSELFPPCLYRHKRTALAIRFPSQDSIDRRLGNLETTRDRDFA QASLGLDVAEKRGGSSWHVHANILYAND*
Bcep176gp20	+	8347	8814	467	***no significant hits***	hypothetical novel protein	-	MAYSLTHSLAMKKVGDRLKELLHAHPNYGERGQSKLSRDT GVPQPTISRILRGGSVPEMETLAALAIPFGVTCEWLLTERGPK FVAELYSPQPHDRAADAPRPLPLHVQNLISEIKRANEAGVSEE VFAAVATLLRAIPAQASCADDGDLPHLQP*
Bcep176gp21	-	8750	9082	332	[*Burkholderia pseudomallei* K96243] gi|52209198|emb|CAH35143.1| hypothetical protein, 92, 1.00E-20: 67% over 51 aa***	hypothetical conserved protein	-	MKRETGVRGLRCRPGDLARIKEAWNPALIGRIVLVKAAHSET EWVVMLLGEPGITLTKNRKRIAASNCVLAYDSALEPIRAVEA GTGDEITAADVEGRRRRHSLPAPVSL*
Bcep176gp22	+	9420	9956	536	***no significant hits***	hypothetical novel protein	-	MSYQLKQFTQDEFEKTRAEWSQYLEKHDMFDLEYKRALDS AEGNREYTRTADNEFAYGVFHDGAQTASAIVSMLHTRRAGR DVGWLKMLQVDLAPEFDEVHVQADLDRLRQVIDIYFAAIVG TVKLGAVHPAKVLKLYGRSTYLLRMLVSVAEKLQQDSANLD AEVSMEGRWLVIRAV*
Bcep176gp23	+	9980	10147	167	***no significant hits***	hypothetical novel protein	-	MNIVHLKAQGSDLRVIAERVADEFQKVANEMPVEDAIRALCF SHLRAQAAARIAA*
Bcep176gp24	+	10210	10416	206	***no significant hits***	hypothetical novel protein	-	MGCFPGQTTAREPDANARDRPLPITRRSIVQRSRRKGARPPG EPNAMRSGWQLSGSPLERPPGRFRDH*
Bcep176gp25	+	10474	10590	116	***no significant hits***	hypothetical novel protein	-	MSPRSRCRLAARQPVHPKSKNPLYSGGQPVIERSLGTA*
Bcep176gp26	+	10590	10697	107	***no significant hits***	hypothetical novel protein	-	MAWLSLSRFARLTSSSSLSPQAASMLPKMIAAAIA*
Bcep176gp27	+	10946	11359	413	***no significant hits***	hypothetical novel protein	-	MDQANWKLLTIESDTRVIEAEYVEGRLAGFVSRAKQHDEIT GFRLRERLGNLLHRLAGFVDGRNVNCSIRVGVIKVERIGISR DVSKVDLKGSTANSASKPVAVGANDASTTNIHFDIDLLGQFV RPGKRIVFYFPA*
Bcep176gp28	+	11549	11623	74	***no significant hits***	hypothetical novel protein	-	MNRCSAPSPWPRAAEVKAREMAAR*
Bcep176gp29	+	11620	11712	92	***no significant hits***	hypothetical novel protein	-	MSCYRIAAPSDEWRGNSSQPRAARNRHGRE*
Bcep176gp30	+	11754	11915	161	***no significant hits***	hypothetical novel protein	+	MRTHRRSTASYAFGRTGGIGSRFPGISVGIGVVGTWWLRPVM LFIGRFLQKTG*
Bcep176gp31	+	11915	11992	77	***no significant hits***	hypothetical novel protein	-	MRLPFLGPAFSLSMKESPTNQRKEK*
Bcep176gp32	+	11989	12165	176	[Bacteriophage P27] gi|18249897|ref|NP_543085.1| hypothetical protein P27p33, 115, 3.00E-08: 53% over 29 aa***	phage P27p33-like	-	MSFKFNLNDAVTIAASGETGTVIGRAEYSNAENGYLIRYKAG DGRATESWWSESALAS*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp33	-	12275	12739	464	***no significant hits***	hypothetical novel protein	+	MLPRGSRRTSCTRCSGKGGGPVGYVHRMQSPNKQSSTQGPP LEALNVLKNLTLDYWYKMLIPLALGLFVLAITVDLRVISNGS LVLLAIGLFMIGLGEWINHPRQEAINLHLNMKITLHRWKPKA WGIALDLAGAVLGGIGFFGIVRAAIRAAA*
Bcep176gp34	+	13022	13162	140	***no significant hits***	hypothetical novel protein	-	MTTDMNTTAPRFTVTLAALRKAGACAARRRPRTAPSSSSCA GSKLS*
Bcep176gp35	+	13173	13256	83	***no significant hits***	hypothetical novel protein	+	MFRLDVLFVIALLIVVYIIAASGSGTC*
Bcep176gp36	+	13277	14023	746	[*Bordetella bronchiseptica* gi|33575614|emb|CAE32692.1| phage-related hypothetical protein, 155, 5.00E-43: 70% over 84 aa***	hypothetical conserved protein	-	MENTPRFTVTLDALRKAGACYEGYNKLVRSLQGQPFTEKDE DRESYIRFRHDAEIPLLDILKSNGLDDAIWTLRCVSGADRDIR LFAVWCARKVEHLMQDQRSKDALDVAERFANGEATEEERA AAWAAARAAGDAAGDAAGAARDAAWVAARAAGDAAGDA AGAARDAAWVAAGAARDAAGAAAWAARDAAGAAAGGAA WAARDAAWAAAGDAAWVAAGAAARDAQTEMFKRMCLGT APWQQGKVAA*
Bcep176gp37	+	14052	14396	344	[Bacteriophage phi1026b] gi|38505435|gb|AAR23204.1| gp53, 115, 2.00E-30: 78% over 89 aa***[Bacteriophage phiE125] gi|17975209|ref|NP_536404.1| hypothetical protein phiE125p48, 138, 2.00E-29: 77% over 88 aa***	phage phiE125 gp47-like; phi1026bgp53-like	-	MTPIDIATLTKVERSILLYAETCCVDAGGLLEGERMNADDMT ALRKFADAGILSFGRIPYHLLASLSGLRQPTHWITLTDDAWQ LAHALRRQRAARGSASRRKVDEVLAEREVT*
Bcep176gp38	+	14444	14719	275	[*Burkholderia cepacia* phage Bcep22] gi|38640345|ref|NP_944268.1| conserved phage protein, 70, 2.00E-04: 54% over 26 aa***	phage Bcep22 gp38-like	-	MDKINDGGPAFPATWANDSDLNAIAPNGQVCPPFSSIPLPGM SLRDYFAAKVMQGALADSNVTLEKAGDADILAKASYRIAD AMLRARGEA*
Bcep176gp39	+	14716	15033	317	[*Burkholderia dolosa* AUO158] gi|84361729|ref|ZP_00986383.1| hypothetical protein BdolA_01001844, 117, 4.00E-14: 66% over 43 aa***	hypothetical conserved protein	-	MNEITTMPELEACGWFVRTKRTDVDPSGLLVADCSAANDRG SMLATLFAASPNMAEILEIVAADADAGTIMLTSGVRLAIDAA LIKAGRKKAPEPVRHFTINGGV*
Bcep176gp40	+	15033	15347	314	[*Burkholderia cepacia* phage Bcep22] gi|38640344|ref|NP_944267.1| hypothetical protein Bcep22p37, 98, 2.00E-14: 44% over 47 aa***	phage Bcep22 gp37-like	-	MALTQEQRVALIVARQYIAEGRDAHLCFALNRVARRYPKLNI AAEGLRAYIQRALSPYTTLEEWIARHELVKPPRLWRIPRTPAE RREARIQWIDWMLDEPKEA*
Bcep176gp41	+	15344	15568	224	[Bacteriophage phi1026b] gi|38505433|gb|AAR23202.1| gp51, 50, 5.00E-07: 48% over 24 aa***	phage phi1026b gp51-like	+	MMRTSLNSLSPVLRDYKQSSVSRAQNDNALLASCEQVHGYV AKGSAVAFAYGVAIGCVWFLCAAYRAGALTWPF*
Bcep176gp42	+	15556	15678	122	***no significant hits***	hypothetical novel protein	+	MAFLKIWIGAAVAIALFLAMQVVLEQRAERFERAPTVQRT*
Bcep176gp43	+	15698	15910	212	***no significant hits***	hypothetical novel protein	-	MCITGPRIARRVFQGRKFRPAFLHRRHRWVLGVTGGRRRTL NRSAAQRPSCYPTLPYASRAPTPDADATN*
Bcep176gp44	+	15924	16397	473	[Bacteriophage 52A] gi|66396299|ref|YP_240658.1| ORF032, gi|23130452|ref|ZP_00112267.1| COG0270: Site-specific DNA methylase, 341, 2.00E-14: 35% over 52 aa***	DNA methylase	-	MSSHQSARMKNDEWLTPPEWLTALGEFDLDPCAPINRPWAT AREHYTIADNGLMREWRGRVWLNPPFGREASKWMRRMAE HGNGVALIPARTETATFFETVWAAAEAVCFVRGRPHFHYADC RRAPFNSGAPICLIAYGRRNAYALLDANLGHVVQV*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp45	+	16416	16766	350	[Bacteriophage phi1026b] gi|38505435|gb|AAR23204.1| gp53, 115, 2.00E-33: 75% over 84 aa***[Bacteriophage phiE125] gi|17975209|ref|NP_536404.1| hypothetical protein phiE125p48, 138, 1.00E-32: 74% over 83 aa***	phage phiE125 gp47-like; phi1026bgp53-like	-	MTTIDINALTKVERSIILYAESCCVDCGGLLEGIRMNADDMTA LRKFADAGILTFGCIPARLLGPLSGLREPTHWITFTEEAWQLA HALRRERAARVTASRRKVDDALAQRATEAA*
Bcep176gp46	+	16763	17947	1184	***no significant hits***	hypothetical novel protein	-	MTPHNEKSPAPAPADEQPASAPIDTPTDREFALSAAICVHPKG VYGDEGGSMCCPITHTRDASFAPSPADERAAFESLKGAWQS MPPFDVFCAGWQAGRAASANETGAEGAEPMAIPAGWKLVP VAPTFEMCRAMEEAIDAGWKDSIVWARGIAAAPQPPAQADA RAGLTDEQRSTLESLARTSTPYEQEVLRSILAAHPGQPVADD AAVRGDERAKRKAFTAVFLDVTTDEAREIVNHPKWSAGSWS HALDDRDAARAELARAAFSQANPGQPEPRAEVTDTEPQWIV NDLGELGVKVGGRFFFLYKGDNIEYGASSSRDGIALHDDGSP MQYRIVGKREFGETCLPVKWVLQGRSEDRYHEKLVYHPGLS FGKPEDGDWKLLPAARTQGGES*
Bcep176gp47	+	17944	18483	539	[*Bacillus* phage SP82] gi|495453|gb|AAA56884.1| endodeoxyribonuclease 168, 158, 4.00E-11: 57 over 34% aa***	endonuclease	-	MKITDDMLTEWRDVVGFEAFYEVSASGQIRNKRNNRIVSGT VYSKGYVVVCLSVGGNVERRLAHKIVAESWIGRRPDGMQID HIDGNRRNNSKYNLRYVTAHQNILATVARGRQAAGTRNGQA RLTSEQVEEIRATRAKGGRYWGGKATCGSLRRDIADHSESRG AQKSPSICPQDRPH*
Bcep176gp48	+	18502	18762	260	***no significant hits***	hypothetical novel protein	-	MIPASEARELAGPTIRERVEALEPLIRAAAEKKQRQIILHDWW ANVGYEGGAAWKEAEKILKEFGYTLEFFYEEQQFVNMYAIV RW*
Bcep176gp49	+	18763	19206	443	***no significant hits***	hypothetical novel protein	-	MAIDTDKMKALAAKLRRYAGSGFVVEAADAIDTLLAALESS KQEADALLSTIAECRDKAHAEGYGESFLYEAIACPDDVPAFI GQTVTELRAALEAAAAYKRDAERYRLLRRGQKWSVVNGIG EMLRAEQLDDAIDAQLSQRQEEGS*
Bcep176gp50	+	19203	19622	419	[*Burkholderia cenocepacia* HI2424] gi|67662640|ref|ZP_00459928.1| hypothetical conserved protein 349, 142, 4.00E-32: 88 over 50% aa***	hypothetical conserved phage protein	-	MKRLTKAEREQVRQMFDGRCAYCGELLSDRWHADHFAPVR RDVKIVKTDRGYRWRSGAPTRPEFDVIDNFMPSCAPCNIDK HAMDLEEWRKKLSRTLDVLNRNYPTYRHARRFGLVAEISVPI VFYFERRAPPSEGEQK*
Bcep176gp51	+	19619	19765	146	***no significant hits***	hypothetical novel protein	-	MSDTDDDWTEPEKEIDDGYARCPHGMPKWDCDWCRESEDI ERVGGTSE*
Bcep176gp52	+	19762	20049	287	[*Polaromonas naphthalenivoran*s gi|84711647|ref|ZP_01019793.1| hypothetical conserved protein 137, 92, 2.00E-07: 41 over 50% aa***	hypothetical conserved protein	-	MTKKQSKHAEPTTDMQVVAPINSQISAVHFQGLRITTPDGQP ATLAVVDMQGRVIESGQTVVRAAWEVAIKAYRNFLSDKGILI VKTISPDKQPT*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp53	+	20206	21888	1682	[*Pseudomonas putida* KT2440] gi|26989745|ref|NP_745170.1| phage recombinase, putative, 541, 5e-137: 54% over 296 aa***	phage recombinase	-	MAKKPKAKVYSYLRFSDPKQAAGSSADRQMEYAARWATEH DMQLDATLTLRDEGLSAFHQRHIKQGALGVFLRAIEDGRIQP GSVLIVEGLDRLSRAEPIQAQAQLAQIINAGITVVTASDGREY NRERLKAQPMDLVYSLLVMIRAHEESDTKSKRVKAAIRRQC EGWVAGTWRGIIRNGKDPHWVRLGEHGKFEHVPERVLAVR TMIDLFLEGHGAIEITRRLTEQNLYVSNAGNYSVHMYRIVRN QALIGEKRISVDGEEFRLDGYYPPILTREEFAELQQTMSERGR RKGKGEIPNIITGLSITVCGYCGRAMTTQNSKARAPKGKSVV RRLSCPMNSFNEGCPIGGSCESEIVERALMRYCSDQFNLSRLL EGDDGTARRTAQLAVARQRASDIEAQIQRVTDALLSDDGKAF AAFTRRARELETQLEEQRREIEALEHQIAASSAHGIPAAAEA WAQLVDGVLALDYDARMKARQLVADTFRKIVVYQRGFAPI DDAAADRWKRSGTIGLMLVTKRGGMRLLNVDRRTGCWQA EDDLDPSLIPSDGLPMLPLDA*
Bcep176gp54	-	21898	22098	200	***no significant hits***	hypothetical novel protein	-	MNAPEIDWKQAPKSARWWAMDADGHAYWYLAPDVAAFTD FWFSESKPAPTFGFVGDWRTSLVERPA*
Bcep176gp55	-	22172	23038	866	[Bacteriophage phi1026b] gi|38505409|gb|AAR23178.1| gp27, 292, 5.00E-67: 48% over 146 aa*** [Bacteriophage phiE125] gi|17484050|gb|AAL40301.1| gp28, 292, 5.00E-67: 48% over 146 aa***	phage phiE125 gp28-like; phi1026bgp27-like	-	MKKIIAALAIPLFLSLAACGGGDDSGPDSSGPAIKLTYSGMPL VTSQRARTMAAATDVSSSASAPGVPANDVQPTISALQDAFK ARGADIAIYPGVINGSKLHDIVMSENNGVGPTRAEIAEANVN ISSWTLVNFQYDDMTGYIDTPEKKAMADQFYKDIRIFAAREY VKGNVVFYANPIMSCLPAKIDSSGYSVATATQSLRDALAGSD DNLGHAIGGISPTPDQMGSDCQTPNSAAQASYVDSIADPLTE QYKTALDTINKCKYDPESIPEEGRWAQCWGIEPVKK*
Bcep176gp56	-	23101	23589	488	[*Burkholderia cepacia* complex phage BcepC6B] gi|47779019|gb|AAT38382.1| gp23, 162, 2.00E-21: 89% over 85 aa***	Rz; phage BcepC6B gp23-like	+	MLKAVLPYLLVALLGAAAGAGVVHLISAREIADMKAAAAQ AQAKAVDAARAEEQRRTAAQSEIAKDANQQRTAALADAFA ARAAAGSLQQRVDQLVAAARHPATSAGSPAAGDALDLLADV LGRVDERAGELAEYADRARIAGQQCERDYDALTIGSSTSRN*
Bcep176gp57	-	23130	23360	230	[*Burkholderia cepacia* complex phage BcepC6B] gi|47779020|gb|AAT38383.1| gp24, 87, 2.30E-02: 80% over 17aa***	Rz1; phage BcepC6B gp24-like	-	MLLLLVLPLAACSSVSTSSSQPPAIPPLPPEARQPATPSICLPTC SAALTSARESWRNTLTVPASPASSASATTTR*
Bcep176gp58	-	23589	24086	497	[Bacteriophage phi1026b] gi|38505406|gb|AAR23175.1| gp24, 163, 2.00E-71: 83% over 134 aa*** [Bacteriophage phiE125] gi|17975186|ref|NP_536381.1| putative lysozyme, 134, 1.00E-64: 115 over 630 aa***	SAR endolysin; phage phiE125 gp25-like; phi1026bgp24-like	+	MADVPKKTLVSVVGAAAAALLFSIVPKFEGLELVARPDPIGII TACNGDTKDVRAGQRFTPDECRARLEQRLIEHAEPVLKCTPS LKGHTYQLAAAVSFAYNVGSGAYCSSTTAKRFNSGDWKGA CRALNEADNGRPQWVTAGGRVLPGLVKRRAEERALCERGL*
Bcep176gp59	-	24079	24291	212	[Bacteriophage phi1026b] gi|38505405|gb|AAR23174.1| gp23, 70, 7.00E-22: 75% over 53 aa*** [Bacteriophage phiE125] gi|17484046|gb|AAL40297.1| gp24, 70, 7.00E-22: 75% over 53 aa***	type II holin; phage phiE125 gp24-like; phi1026bgp23-like	+	MRASPTEVASYVGSVTAVASSLTLTDIGVIVGILTAIATFGLNF FMWRKDRREQRESDMRIMEMEKHDG*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp60	-	24332	25048	716	[Bacteriophage phi1026b] gi|38505404|gb|AAR23173.1| gp22, 244, 4.00E-51: 43% over 106 aa*** [Burkholderia thailandensis E264] gi|83717361|ref|YP_439261.1| gp22, 244, 1.00E-91: 67% over 160 aa***	phage phiE125 gp23-like; phi1026bgp22-like	-	MPAGFQAFTDTGLYQIDGRTPNYQMVQAMVGQAVNSDLHL AYNDANREFRASMPNVTFTFNANAGPMYGVYASGGTGITL WAAKRSDLVYTLTFVTEQPCTVYLFLFDQVPPAAGNFGMQV FDAGGVLIADSSRPFLRVLDVIYDEYIPGTGWAVIGRPSPPWQ SRAYAAPVIASAIYSVRKIWWNDPPGVQLTSIRVTGNVVSWG TMIHGDGGGNNFVGFREQFHSRFMVLDGTGIV*
Bcep176gp61	-	25048	25356	308	[Bacteriophage phi1026b] gi|38505403|gb|AAR23172.1| gp21, 104, 3.00E-24: 55% over 57 aa*** [Bacteriophage phiE125] gi|17975183|ref|NP_536378.1| hypothetical protein phiE125p22, 104, 3.00E-24: 55% over 57 aa***	phage phiE125 gp22-like; phi1026bgp21-like	-	MVAGLQIFDGAGRLLLDAKSRAGRVVGIVHTGGVDGSVAAD MSGGEPFWAFMPEQLFYRVSGAEPSPVVSINAGGISWSYSPN DSGSNAYTRVPGWIVFGVY*
Bcep176gp62	-	25356	28658	3302	[Bacteriophage phi1026b] gi|38505402|gb|AAR23171.1| gp20, 1101, 0: 79% over 878 aa*** [Bacteriophage phiE125] gi|17975182|ref|NP_536377.1| putative tail tip fiber protein, 1101, 0: 79% over 880 aa*** [*Burkholderia thailandensis* E264] gi|83716294|ref|YP_439259.1| host specificity protein J, 1101, 0: 80% over 883 aa***	host specificity protein J; phage phiE125gp21-like; phi1026bgp20-like	-	MKKLYAESGLKRISGAKGGGGGGGGGESPDSLHSTARAKVL DIIGEGPIVGLVNGMQSVYLDGTPIQNADGSVNFQNYTVDVR TGTLDQDYMPGFPAVERETAVGIPLTSDAPWVRQVQNTQLT/ VRIRFGVPALQKSDPSSGVFGYRVEYAIDLSVDGGSYAEVLR GAFDGKTTSLYERSHRIELPSAKSGWLVRVRRITPNAHSSLIA DTVNIEAITEIIDRKLRYPMTALVGMTFDARSFSQVPVRSYHV RGLIVRVPSNYDPETRTYSGTWDGTFKPAWTNNPAWIFYDLI LNERYGLGKTVDASMIDKWGLYEIARYCDVMVSDGKGGVE PRFACNCVIQSAADAFKVLQDIASVFRGIAYWGPGAVVASAD MPSDPVYVYTAANVIDGSFRYVGSERKTRYTVALVSYNDPA NQYKQAVEYVPDEDGIARYGVIKTQVTAFGCTSQAQAHRVG QWILLTSRYEAGTVSFQVGMDGVLVGPGQVIAIADPRKAGR RIGGRIRSAAGAVVVLDKAPAVAPGDRFTAILPSGVAQSRAVI SVDGDTLTLVDRFDADPVSGAVWMLESDALTAQLYRVVSVC ESDEEGQIAYTITATRHEPGKYAAIDDGAQIQQRPVTVVPPSV QAPPSNVRLTTYSVIDQGIAKTNMVIAWDAADKAVTYLPEW RKDNGDWVAVAQTGGLQVEIPGIYQGRYVARVRAQNVMNV TSLPAVSAETMLMGKTTPPPSVVSLKASGVVYGINLEWKFPG DGSAGDTQRTEIWYSRTPSRDDATKFSDFAYPQASTSYQGLA VGQVFYFWARLVDTSGNVGPWYPADGPGVQGQPTTDEAAY EDYFRGQIGKSSLGHDLLEPIDSITPPMAGDAEEYAGDATMY AGVWSLQSAIAEGDMAVAQKVETVAAQMNSATATLSAAVQ EAKARVDADSAMAQQITTVQAKANENAAAVQTVAQSYADL NGRVSASYQIKTQITSDGRTYIAGIGVGVDNNNGIIESQVLVS AQRFAVVDPNNGGSMIVPFVVQGGQVFIRQALIGAGWITNA MIGSYIQSDNYIAGRQGWRLDKSGWFEINAADGSGNRLVLDC

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp63	-	28655	29218	563	[Bacteriophage phi1026b] gi|38505401|gb|AAR23170.1| gp19, 194, 5.00E-71: 72% over 140 aa*** [Bacteriophage phiE125] gi|17975181|ref|NP_536376.1| putative tail component protein, 194, 2.00E-71: 72% over 140 aa*** [*Burkholderia thailandensis* E264] gi|83717443|ref|YP_439258.1| Bacteriophage lambda tail assembly protein I, 194, 4.00E-70: 71% over 138 aa***	tail assembly protein I; phage phiE125 gp20-like; phi1026bgp19-like	-	MSEKLREVRLYGIAGTRFGRVHRLAVSSTAEAVRALSVLLPG FRKFLLEARDNGLTFAVFNGRRNLSQDDLTAPVGDEAIRIAPV IIGSKSGGLFQTILGAALAVVGFVFEQPSLIAMGASMMLGGV AQMLSPHQPGLAGVTDNGTSYYFNGPVNSAAQGEPVPLIYG DTITGSKVASSGIYAEDQQ*
Bcep176gp64	-	29215	29967	752	[Bacteriophage phi1026b] gi|38505400|gb|AAR23169.1| gp18, 250, 2e-113: 214 over 76% aa*** [Bacteriophage phiE125] gi|17975180|ref|NP_536375.1| putative tail component protein, 257, 2e-114: 215 over 76% aa*** [*Pseudomonas fluorescens* Pf-5] gi|70731109|ref|YP_260850.1| prophage LambdaSo, tail assembly protein K, 254, putative, 1.00E-58: 46% over 119 aa***	tail assembly protein K; phage phiE125 gp19-like; phi1026bgp18-like	-	MDERIKQAIADHALAEYPRECCGLIVRTAAGDVYLPGRNVA PTPTDQFALAPEDYADAEDMGEIVAMVHSHPNGTAQPSMAD RTVCERAGIPQWVIVSLGVQADGSIGVDDWNEFGPSGYVAPL YGREFLHGVLDCYALIRDWYLAERGITLPDFERKDGWWAD GHSNLYIAHYQDAGFLDVGRDVALEPGDVLLMQIRSKNGVP NHAGVYLGDGMFAHHMYGRLSGKAVWGSMWRDSCTTVLR YVGGAN*
Bcep176gp65	-	30016	30699	683	[Bacteriophage phiE125] gi|17975179|ref|NP_536374.1| putative minor tail protein, 227, 2e-107: 202 over 78% aa*** [Bacteriophage phi1026b] gi|38505399|gb|AAR23168.1| gp17, 227, 7e-112: 205 over 81% aa*** [*Burkholderia thailandensis* E264] gi|83716138|ref|YP_439256.1| phage minor tail protein L, 227, 3e-112: 205 over 81% aa***	tail assembly protein L; phage phiE125 gp18-like; phi1026bgp17-like	-	MAISADVQSLEPGHRIELFEVDCTAIGGDVLRFHGHLQSTSIV WQGQEYKPWPIQAAGFERTSDARQPAPTLTVGDINGTITAMC VALEDLVGAKVFRRRTLAKYLDAVNFPDGNPTADPNEQWPV EQWRIEQKSDEQPGVQVEFTLSSPLDFGGQQVPARQIVGTCQ WRYRGPECGYTAMVFFDKNDNPVSDPALDRCSQRISGCERR FGVNNPLPYGGFLCDTLA*
Bcep176gp66	-	30742	31749	1007	[Bacteriophage phi1026b] gi|38505398|gb|AAR23167.1| gp16, 462, 2.00E-41: 50% over 83 aa*** [Bacteriophage phiE125] gi|17975178|ref|NP_536373.1| hypothetical protein phiE125p17, 462, 5.00E-39: 49% over 81 aa***	phage phiE125 gp17-like; phi1026bgp16-like	-	MSTLQKINLGTPPRGDDGDSTRVGFAKMNANVDVLSMLVAL GYALITDNTTLTPDQVGTRYGINIADPGKSITIPLASSVPVNAC LQFFNVGSAVTVATQGNDGTQIRALNKGDWATYISDGVKYW HVAERGKMLSDEIVGGNLTVGGRVLLGGASDDGSTALQVKG AASIGKFLSRPTFADNVPWDSGNFDPASKLSVVGRKDGIAPA AGNVGEILTSTNSTIVLPSNRTAEGTTTLLLTAGEWDVQGNM FNYNANGVTLTMAFVSVATSSASIVGGQFAGIDGLQTMGWV TFVTPMVRFRFAAPTTVWLNAQAAANVNVPCWGQLTARRV AA*
Bcep176gp67	-	31758	32096	338	[Bacteriophage phi1026b] gi|38505397|gb|AAR23166.1| gp15, 112, 7.00E-45: 75% over 84 aa*** [Bacteriophage phiE125] gi|17484038|gb|AAL40289.1| gp16, 112, 7.00E-45: 75% over 84 aa*** [*Pseudomonas fluorescens* Pf-5] gi|70731112|ref|YP_260853.1| prophage LambdaSo, minor tail protein M, putative, 125, 3.00E-21: 48 over 256 aa***	tail assembly protein M; phage phiE125 gp16-like; phi1026bgp15-like	-	MTDTFNWSPTVEGFGGDTTLRVRKAGFGDGYTQRAADGLN NRTATYNLRFVGRADMIAAILAFLDAHAGSVSFFWTPPLRPQ GRFVCEKYSEPVKNGEVYTITAQFEETFAP*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp68	-	32096	36238	4142	[Bacteriophage phi1026b] gi|38505396|gb|AAR23165.1| gp14, 1354, 0: 69% over 952 aa*** [Bacteriophage phiE125] gi|17975176|ref|NP_536371.1| putative tail length tape measure protein, 1354, 0: 943 over 3460 aa***	tape measure protein; phage phiE125 gp15-like; phi1026bgp14-like	-	MATSLRELIVSVTANTTEYDRRMRGLASTAGGYFNAVRDGG RAADAAFASNAASVQVTVRAIEAARGSLRGYAEAAAAAFGI HQLVDYADEWTNLSNRLKIVTRDQVDFAVAQDDVLRIAQAT RQPLDATAELYQRIANNTSHLGLSIKQVGPLVETISKAVALSG VSADTARLGIVQLGQAFASGQLRGQDLKSVLEELPGVADAIA RGMGRGTSELKALAEDGKLTVENLIDALRNAGASTDALFGK VDVTVGQAMTRLQTEIVAYVGRANEATGASAKLSQGIIYVAI HLDDIVKISASLAAGRLVTYLLQSGIAARQAAIDWNTKRLAL AAATKEESAAALVSLEKARSDRDAAAAKLQNAQAAEVAAQ AELAGMRAMRESLAMQSALIAGSIRYTDAKLAEARAIEASA VAQVATARSNLANSQEIGTRIAGTPYAAVIARETAAAQGELEI AEASLALAQQRRVALEAAAAKGTIDQTRYAAALAETEKGLA AAEREVAIATQARERAERGATAATAGLAAATERAAVAQTAA. RAGSLMRTVGSGLLSVVGGLPGILTTVGTVALGAAVNWLVF RDHASSATSSLIDMQAPLDQVIDKYRQLTPLLQEVERQRVKQ AQAAATSDVADAYARLASRASQSVIVPTFGDGAPIVTDEDQA ALDRFLEGLKRVKAENLGVEEKSRELARLVDVFVAATRGGD DLRTELVQDASAIDTAGAAADRGARTLAAMDAAARGAADG IRLLTEENNFFAGGMAAEAWNKYVEKLKEASDVIGMTAQQR AEYEAKTKGANAAEARQAGLIAGRADAYKSLEKAIQDKDA KAEAGARRNIDNLTRELALMNQQMEVAKQLSGYQNAIKNG DFSQFGIGKETPEKIALEVAGNVARARAAALGQKAYDDTMA QSAAQVARIGVNAPALAHRTRAGGARSEPESRRLLENIEQRI AQLRVEAVATDKLTQSEKDRIGFDQKLTDLAAKRTKLSDGDI SLIRDQASIRAAYDRAVQLEKEVRYHEAINKLKERSAQIDAEL ADYASERQREVARELAALSMGDNARELNQAMSRVSDEFRRR
Bcep176gp69'	-	36251	36990	739	[Bacteriophage phiE125] gi|17975174|ref|NP_536369.1| putative tail assembly chaperone protein, 156, 1.00E-43: 85 over 449 aa*** [Bacteriophage phi1026b] gi|38505394|gb|AAR23163.1| gp12, 154, 2.00E-42: 57% over 84 aa***	alternative C-terminus for gp69; lambda T equivalent; phiE125gp14/phi1026bgp12-like	-	MTNKVKPGALRAAILNPLAGWRHEFMPMPEWDGVTVAVRE PLLEDRAFWLDPLREAAGVEPDDDEQTARSKYSRVSADAHK LAYARLFVRVLYVETAQGWRREFGDSDAQAVAEAYGPAHDI VVTKAIELGGLAADREEAGKKAFRRNPDLRLEMTLALRLGR TLAELRAQMSAAEFALWQAFDAESPISEERYDLHAAMVASA VFQAQGAKVKVADMMPKWGGESEDAQEIEDDPFFVGLMSL AK*
Bcep176gp69	-	36526	36990	464	[Bacteriophage phiE125] gi|17975174|ref|NP_536369.1| putative tail assembly chaperone protein, 156, 7.00E-41: 58% over 79 aa*** [Bacteriophage phi1026b] gi|38505394|gb|AAR23163.1| gp12, 154, 9.00E-40: 94 over 57% aa***	lambda G equivalent; tail assembly chaparone; phagephiE125 gp13/phi1026bgp12-like	-	MTNKVKPGALRAAILNPLAGWRHEFMPMPEWDGVTVAVRE PLLEDRAFWLDPLREAAGVEPDDDEQTARSKYSRVSADAHK LAYARLFVRVLYVETAQGWRREFGDSDAQAVAEAYGPAHDI VVTKAIELGGLAADREEAGKKPSAETPTSDSK*
Bcep176gp70	-	37019	37483	464	[Bacteriophage phi1026b] gi|38505393|gb|AAR23162.1| gp11, 152, 4.00E-62: 80% over 117 aa*** [Bacteriophage phiE125] gi|17975173|ref|NP_536368.1| putative major tail subunit protein, 152, 4.00E-62: 80% over 117 aa***	tail component; phage phiE125 gp12-like; phi1026bgp11-like	-	MAEKSKRTKAQGTKVEVSKTVSTDLDDNTLVFVDLNTTGK TIQWQGGQSSEIDATTLASEEKEYELGLPDPGEFSVDGNYSS DDEGQSLLRTARASGEKHVFRVTFADQSQFLFVGMVRQYTW SAAVDGIVTSTYSVRVSGAPKLVPPPAPPAG*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp71	-	37540	37884	344	[Bacteriophage phiE125] gi|17975172|ref|NP_536367.1| hypothetical protein phiE125p11, 115, 4.00E-33: 60% over 70 aa*** [Bacteriophage phi1026b] gi|38505392|gb|AAR23161.1| gp10, 115, 5.00E-33: 60% over 70 aa***	phage phiE125 gp11-like; phi1026bgp10-like	-	MSALTIRDAIGIVGGAKGYVSVASSAAQSPYYVVSRVSGTRD MALGGATGGKSGMFQIDVYAKTYTEADSLADQIIDRVESTG MFSVGGVSDLPDDYSSDTGVFRVSLEISVQF*
Bcep176gp72	-	37881	38303	422	[Bacteriophage phi1026b] gi|38505391|gb|AAR23160.1| gp9, 140, 6.00E-61: 117 over 598 aa*** [Bacteriophage phiE125] gi|17975171|ref|NP_536366.1| hypothetical protein phiE125p10, 140, 5.00E-60: 83% over 115 aa***	phage phiE125 gp10-like; phi1026bgp9-like	-	MSSVQILGLADLQADFLKLAKAQSTKALRRATVAGANVIRD EARARAPKKTGKLKRNIVTAALKQKDSPGIATAGVRVRTKG KADSPNNAFYWRFVELGTQFMKAEPFMRPAFDASIAQAEGA IRTEIARAIDQVVGGGL*
Bcep176gp73	-	38296	38625	329	[Bacteriophage phi1026b] gi|38505390|gb|AAR23159.1| gp8, 109, 4.00E-39: 77 over 410 aa*** [Bacteriophage phiE125] gi|17975170|ref|NP_536365.1| hypothetical protein phiE125p09, 108, 9.00E-32: 62% over 67 aa*** [*Burkholderia thailandensis* E264] gi|83717783|ref|YP_439247.1| phage head-tail adaptor, putative, 109, 1E-35: 64% over 70aa***	head-tail adaptor; phage phiE125 gp9-like; phi1026bgp8-like	-	MLRSSDLTEFIVIERKGGRTNENGEPLPDDWVTHDEVWASV RFVSGKEHVISGAVRSSAIASIRIRFREDIDSEMRIRYGDQLYD IVAVLPNRRKGSLDLPVKVGEKYV*
Bcep176gp74	-	38628	38969	341	[Bacteriophage phi1026b] gi|38505389|gb|AAR23158.1| gp7, 120, 2.00E-19: 52 over 239 aa***	phage phi1026b gp7-like	-	MALVELKLALGFVRANAGVEDDVVQMLLDAATQSAVDYLN RQVFETEDAMTTAIEAGTAGQNPMVVNAAIRAAILKITAELY ANREDTAFGPITELPLNARALLRPHRIIPGV*
Bcep176gp75	-	38973	39137	164	***no significant hits***	hypothetical novel protein	-	MATLIKPFKGVKNGEIYPTEFEAGDECPEELEDGARAVGALE GAEEKKPTAPKK*
Bcep176gp76	-	39211	40488	1277	[*Burkholderia thailandensis* E264] gi|83716436|ref|YP_439244.1| phage major capsid protein, HK97 family, 441, 0: 85% over 363 aa***	major capsid protein	-	MSKKLLIAVLTAALTGPVGAVPRGIISVRAEGPTEVKALIENL QKAFHDFKAEHTKQLDAVKAGLPTSDALAKVDKVSADLEA LQAAVDEANIKIAAAQMGANGVKPLRDPEYTEAFKAHVKR GDVQAALNKGEDSEGGYLTPIEWDRTITNKLVLISPMRQLCR VQPVSKAGFSKLFNMGGTTSGWVGEASQRPQTNAATFQPLS FASGEIYANPAATQQILDDAEIDLESWLATEVQTEFAKQEGKA FLAGDGTNKPNGLLTYIAGGANAAKHPFGAIEVVNSGAAAD ITSDGIIDLVYDLPSAFTGNARFAMNRNTQRQVRKLKDGQG NYLWQPSYVAGQPATLAGYPVTEVPDMPDVAANSTPILFGDI QQTYLIIDRIGVRVLRDPYTAKPYVLFYTTKRVGGGLLNPEP MRAMKVAASE*

Table 2. Continued

Bcep176 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM Domain	AA sequence
Bcep176gp77	-	40493	41335	842	[B*urkholderia thailandensis* E264] gi|83718056|ref|YP_439243.1| ClpP protease, 280, 4.00E-114: 268 over 91% aa***	Clp-related prohead protease	-	MSLLQLPEIRADHRLGAAQFDVRPDALERWNPDVRAANAD DAAAISIYDSIGDNWEGTGVTAKRISAALRNIGAREVTVNVN SPGGDFFEGVAIYNLLREHKAKVTVNVMGLAASAASVIAMA GDEILMGDGAFLMIHNAWAVAIGNRHDMAAAAATLAPFDA AMAKLYAQRAGISEETAAAMMDKETWIGADQAVKDGFATG KLDSAPIAMDTNTSSGRKAMALIEASMARAGYSRSMRRDAL KALFDGKPGATAGNATPGAGEDVAASLHTLINALKG*
Bcep176gp78	-	41313	42608	1295	[*Burkholderia thailandensis* E264] gi|83716795|ref|YP_439242.1| head portal protein, 460, 0: 90% over 391 aa*** [Bacteriophage phi1026b] gi|38505385|gb|AAR23154.1| gp3, 437, 3e-28: 30% over 110aa*** [Bacteriophage phiE125], gi|17975165|ref|NP_536360.1| putative portal protein, 419, 5e-29: 31% over 107aa***	portal protein; phage phiE125 gp4-like; phi1026bgp3-like	-	MGLFDFIRREKQPEAQARPHVEPSFQASTPTTSIPGETFEGLD DPRLKEYIRRGELNGGTGRETRALRNMAVLRCVTLISGTIGM LPMNLISSDDSKQVLTDDPAHRLLKYKPNDWQTPMEFKSLM QLRALLDGESMARIVWSGNRPIRLIPMDRGSAKGRLTSTWQI VYDYTTPTGDKIELPAREVFHLRDLSIDGVSGVSRVKLSGNA LELAEQAERAASRTFRTGVMAGGAIEVPKELSDNAYGRMKA SVQENHTGSENAGSWMLLEEGATAKQFSNTAASAQQIENRN HQIEEVARMYGVPRPLLMMDDTSWGSGIEQLAIFFIQYGLSH WFVSWEQAAARAFLPEKMLGQRQFKFNEGALLRGTLNDQA AFFSKALGAGGQSPWMKQNEVREMLDLPRADDPVADQLRN PMTQKQKGSGDEPPATT*
Bcep176gp79	-	42616	44328	1712	[*Burkholderia thailandensis* gi|83717767|ref|YP_439241.1| phage terminase, large subunit, 580, putative 2418, 0: 528 over 84% aa***	TerL	-	MEWTTACPDWERRLIARESIIPPPIFPDQAEQAVAIFKELRVCD LPGKPTFGECSEQWVFDFVAAIFGAYDAETGKQLIREFFLLIS KKNTKSTIAAGIMLTAVILCWREEEEHLILAPTKEVADNSFKP AAGMIRADAELSELFHVQDHIRTITHRVSRASLKVVAADTDT VSGKKSGKILIDEHWVFGKRANAEAMFMEATGGQVSRDEG WVIILTTQSDEPPAGVFKEKLQYYRDVRDGKIVDRKSLGVLY EFPEEMVKSKAYLDPANYYITNPNLGRSVSAEWLEDQLNKIR TKTDGSFQQFIAKHLNVEIGMNLRSDRWAGADFWINAGLPE RVTLLDLIEQCEVIAAGIDGGGLDDLLGLSAVGRVRGGRNWL AWAHAWAHPSVLERRKEIAPALHDFEKAGDLTIVSRIGEDVV QAAEYVARIERAGLLYKAGVDPAGIGAILDALAAARVPEDK VIGISQGWKLSGAIKTTERRIAAASGQRIEGDEAPDGALYHG GQPLLTWAVGNARVVPVGNAVNITKQVSGTAKIDPLMALFN AVSLMALNPPAQGQSVYESRGIRFL*
Bcep176gp80	-	44328	44765	437	[*Burkholderia thailandensis* E264] gi|83716048|ref|YP_439240.1| hypothetical protein BTH_II1043, 161, 3.00E-4182% over 120 aa***	TerS	-	MAGKSNKDAAIAAGYSPATASAAGSRLVKDKDVALYIAANR VQKGSNAGDQSQQPAPAQKPDGFDLEAMTNFTDPKAFLIAA MNDSRTEPKLRIDAAKALMPFVHKRLGEGGKKEQRDEAAK KAASRFAPAAPPRLIANGGKKVE*

by its 4.00E-39 similarity to Bacteriophage phi1026b gi|38505390| gp8, and Bacteriophage phiE125 gi|17975170| hypothetical protein phiE125p09 (9.00E-32).

Tail assembly cassette. Tail assembly is only accomplished with the help of a concert of proteins, some chaperoning and some present in the final structure. Several of these tail-related proteins are grouped together: tail component [70], a homolog to Bacteriophage phiE125 gi|17975173| putative major tail subunit protein by an expect of 4.00E-62; tail assembly chaperone [69] similar to Bacteriophage phiE125 gi|17975174| putative tail assembly chaperone protein by a degree of 1.00E-43; tape measure protein [68] determined by expect of 0 to Bacteriophage phiE125 gi|17975176| putative tail length tape measure protein. Tail assembly proteins M, L, K, I and J [67, 65-62] annotated as such due to their respective similarities to *Pseudomonas fluorescens* Pf-5 gi|70731112| prophage LambdaSo minor tail protein M with an expect of 3.00E-21, *Burkholderia thailandensis* E264 gi|83716138| phage minor tail protein L with a strong expect value (3E-112), *Pseudomonas fluorescens* Pf-5 gi|70731109| prophage LambdaSo tail assembly protein K (1.00E-58), *Burkholderia thailandensis* E264 gi|83717443 bacteriophage lambda tail assembly protein I (4.00E-70), *Burkholderia thailandensis* E264 gi|83716294| host specificity protein J with (0).

Terminase cassette. Terminases are enzymes that can recognize phage DNA to be packaged and cut in a way characteristic of that virus (5). TerL and TerS [79, 80] of Bcep176 were most related to *Burkholderia thailandensis* E264 gi|83717767| phage terminase, large subunit with a convincing expect value of 0 and TerS, by position in the genome with respect to TerL and similarity to *Burkholderia thailandensis* E264 gi|83716048| hypothetical protein BTH_II1043 by 3.00E-41.

Lysis cassette. Lysis gene products are involved in timing and creating the pores in the host cell wall that ultimately cause lysis of the host. A distinct lysis cassette follows the tail proteins: type II holin [59], whose functionality was determined through a similarity of 7.00E-22 to Bacteriophage phi1026b gi|38505405| gp23; SAR endolysin [58], annotated by its peculiar sequence and high similarity to several lysozymes including gp24 of Bacteriophage phi1026b gi|38505406| by a similarity degree of 2.00E-

71; and Rz/Rz1 [56/57] – Rz was detected by similarities to other proteins such as *Burkholderia cepacia* complex phage BcepC6B gi|47779019| gp23 with 2.00E-21 and Rz1 was determined due to its characteristic placement within Rz.

Morons. It seems that the only moron identifiable in Bcep176 through database comparisons is the DNA methylase [44] related to *Magnetospirillum magneticum* AMB-1 gi|82945997| site-specific DNA methylase with an expect of 2.00E-10. It's name suggests that it helps protect the phage DNA from detection by its host.

Comparison of Bcep176, phiE125 and phi1026b. 29 of 81 proteins in Bcep176 have homologs in phiE125 and phi1026b. (See Table 3)

In Bcep176, 16 hypothetical conserved proteins [03, 04, 05, 08, 11, 14, 37, 41, 45, 55, 60, 61, 66, 71, 72, 74] correspond to 16 in phi1026b [gp76, 75, 74, 71, 72, 65, 53, 51, 53, 27, 22, 21, 16, 10, 9, 7] and 13 in phiE125 [gp65, 64, 63, none, 60, 55, 47, none, 47, 28, 23, 22, 17, 11, 10, none] respectively with degrees of similarity ranging from 3.00E-05 (Bcep176gp03) to 5.00E-67 (Bcep176gp55). The clustering of these sequentially organized hypothetical conserved proteins indicates that they must have been transferred as modules between phages.

Two particular genes of interest in the lysis cassette of Bcep176 also had shared similarities with homologs in phi1026b and phiE125: the SAR endolysin [58] bore close resemblance to phi1026bgp24 (2.00E-71) and phiE125gp25 (1.00E-64); the type II holin [59] also had a marked similarity to phi1026bgp23 (4.00E-51) and phiE125gp24 (7.00E-22). Since the lysis cassette functions as a whole, it is safe to say that these two proteins involved in lysis can be considered as shared modules.

Nearly all proteins contained within the Bcep176 tail assembly cassette bore close resemblance to homologous proteins in phi E125 and phi1026b that in turn are homologs of lambda tail proteins (M, L, K, I, J, G). The host specificity protein J [62] of Bcep176 had a perfect degree of similarity of 0 to homologs phiE1026bgp20 and phiE125gp21 J protein homologs. Tail assembly protein I [63] also bore a high degree of resemblance to phi1026bgp19 (5.00E-71) and phiE125gp20 (2.00E-71). Continuing the trend, Bcep176gp64, annotated as tail assembly protein K, shared a very close

resemblance to phi1026bgp18 (2.00E-113) and phiE125gp19 (2.00E-114). The annotated protein L from Bcep176gp65 follows suit with expect values of 7.00E-112 for

TABLE 3. Bcep176 proteins to phi1026b and phiE125 mapped proteins

Protein in Bcep176	Protein in phi1026b/gi (gp)	protein in phiE125/gi (gp)	encoded protein
Bcep176gp03	p76/gi|38707966(gp76)	p66/gi|17975227(gp65)	hypothetical conserved protein
Bcep176gp04	p75/gi|38707965(gp75)	p65/gi|17975226(gp64)	hypothetical conserved protein
Bcep176gp05	p74/gi|38505456(gp74)	pp64/gi|17975225(gp63)	hypothetical conserved protein
Bcep176gp08	p72/gi|38505454(gp72)	none	hypothetical conserved protein
Bcep176gp11	p71/gi|38505453(gp71)	p61/gi|17975222(gp60)	hypothetical conserved protein protein
Bcep176gp14	p65/gi|38505447(gp65)	p56/gi|17975217(gp55)	hypothetical conserved phage protein protein
Bcep176gp37	p53/gi|38505435(gp53)	p48/gi|17975209(gp47)	hypothetical conserved protein
Bcep176gp41	p51/gi|38505433(gp51)	none	hypothetical conserved protein
Bcep176gp45	p53/gi|38505435(gp53)	p48/gi|17975209(gp47)	hypothetical conserved protein
Bcep176gp55	p27/gi|38505409(gp27)	p28/gi|17975189(gp28)	hypothetical conserved protein
Bcep176gp58	p24/gi|38505406(gp24)	p25/gi|17975186(gp25)	SAR endolysin
Bcep176gp59	p23/gi|38505405(gp23)	p24gi|/17975185(gp24)	type II holin
Bcep176gp60	p22/gi|38505404(gp22)	p23/gi|17975184(gp23)	hypothetical conserved protein
Bcep176gp61	p21/gi|38505403(gp21)	p22/gi|17975183(gp22)	hypothetical conserved protein
Bcep176gp62	p20/gi|38505402(gp20)	p21/gi|17975182(gp21)	host specificity J protein
Bcep176gp63	p19/gi|38505401(gp19)	p20/gi|17975181(gp20)	tail assembly I protein
Bcep176gp64	p18/gi|38505400(gp18)	p19/gi|17975180(gp19)	tail assembly protein K
Bcep176gp65	p17/gi|38505399(gp17)	p18/gi|17975179(gp18)	tail assembly protein L
Bcep176gp66	p16/gi|38505398(gp16)	p17/gi|17975178(gp17)	hypothetical conserved protein
Bcep176gp67	p15/gi|38505397(gp15)	p16/gi|17975177(gp16)	tail assembly protein M
Bcep176gp68	p14/gi|38505396(gp14)	p15/gi|17484037(gp15)	tape measure protein
Bcep176gp69	p12/gi|38505394(gp12)	p13/gi|17975174(gp13)	lambda G-like tail assembly chaperone
Bcep176gp69'	p12/gi|38505394(gp12)	p14/gi|17975175(gp14)	alternative C-terminus to Bcep176gp69
Bcep176gp70	p11/gi|38505393(gp11)	p12/gi|17975173(gp12)	tail component
Bcep176gp71	p10/gi|38505392(gp10)	p11/gi|17975172(gp11)	hypothetical conserved protein
Bcep176gp72	p9/gi|38505391(gp9)	p10/gi|17975171(gp10)	hypotheical conserved
Bcep176gp73	p8/gi|38505390(gp8)	p09/gi|17975170(gp9)	head-tail adaptor
Bcep176gp74	p7/gi|38505389(gp7)	none	hypothetical conserved protein
Bcep176gp78	p3/gi|38505385(gp3)	p4/gi|17975165(gp4)	portal protein

phi1026bgp17 and 2.00E-102 for phiE125gp18. Tail assembly protein M (Bcep176gp67) adds to the trend of high similarity with a degree of identity of 7.00E-45 with both phi1026bgp15 and phiE125gp16. The tape measure protein (Bcep176gp68) follows the J, I, M, L, and K homologs with a degree of similarity of 0 to proteins phi1026bgp14 and phiE125gp15. Next are proteins Bcep176gp69 and Bcep176gp69', annotated as a lambda G equivalent tail assembly protein and alternative C-terminus to gp69. Due to their peculiar gene product-alternative terminus pairing, both proteins bear differing degrees of resemblance to the same two gene products phi1026bgp12 (9.00E-40 and 2.00E-42, respectively) and phiE125gp13 (7.00E-41 and 1.00E-433, respectively). Last in the cassette, Bcep176gp70 encodes for a tail component that has a homolog in phi1026bgp11 phiE125gp12 with the same degree of similarity (4.00E-62). The striking resemblances between the tail assembly cassettes of Bcep176 and phiE125 and phi1026b indicate that modular exchange must have taken place at some point between these phages.

The remaining annotated proteins that have a putative function by similarity to counterparts in phiE125 and phi1026b are Bcep176gp73 (head-tail adaptor) and Bcep176gp78 (portal protein). The head-tail adaptor [73] bears a strong similarity to phi1026bgp8 (6.00E-61) and to phiE125gp9 (9.00E-32). The portal protein [78] has a close resemblance to phi1026bgp3 (3.00E-28) and phiE125gp4 (5.00E-29). This small but strong example of clustering also supports a modular exchange of head-to-tail proteins.

Similarities between phi1026b and phiE125 proteins with respect to Bcep176 structural proteins indicate a modular exchange easily observable in the area of tail and tail-to-capsid proteins. (See Fig. 7)

Bglu421 analysis. *Genome structure.* The initial assembly of Bglu421 has 50,817 bps of unique coding sequence (Table 4). At this time, genomic ternmini cannot be determined.

Genome overview. Bglu421 is predicted to encode 70 proteins, 20 of which have no homologs in the database (Table 5). Contig A is 5,571 bps long, contains

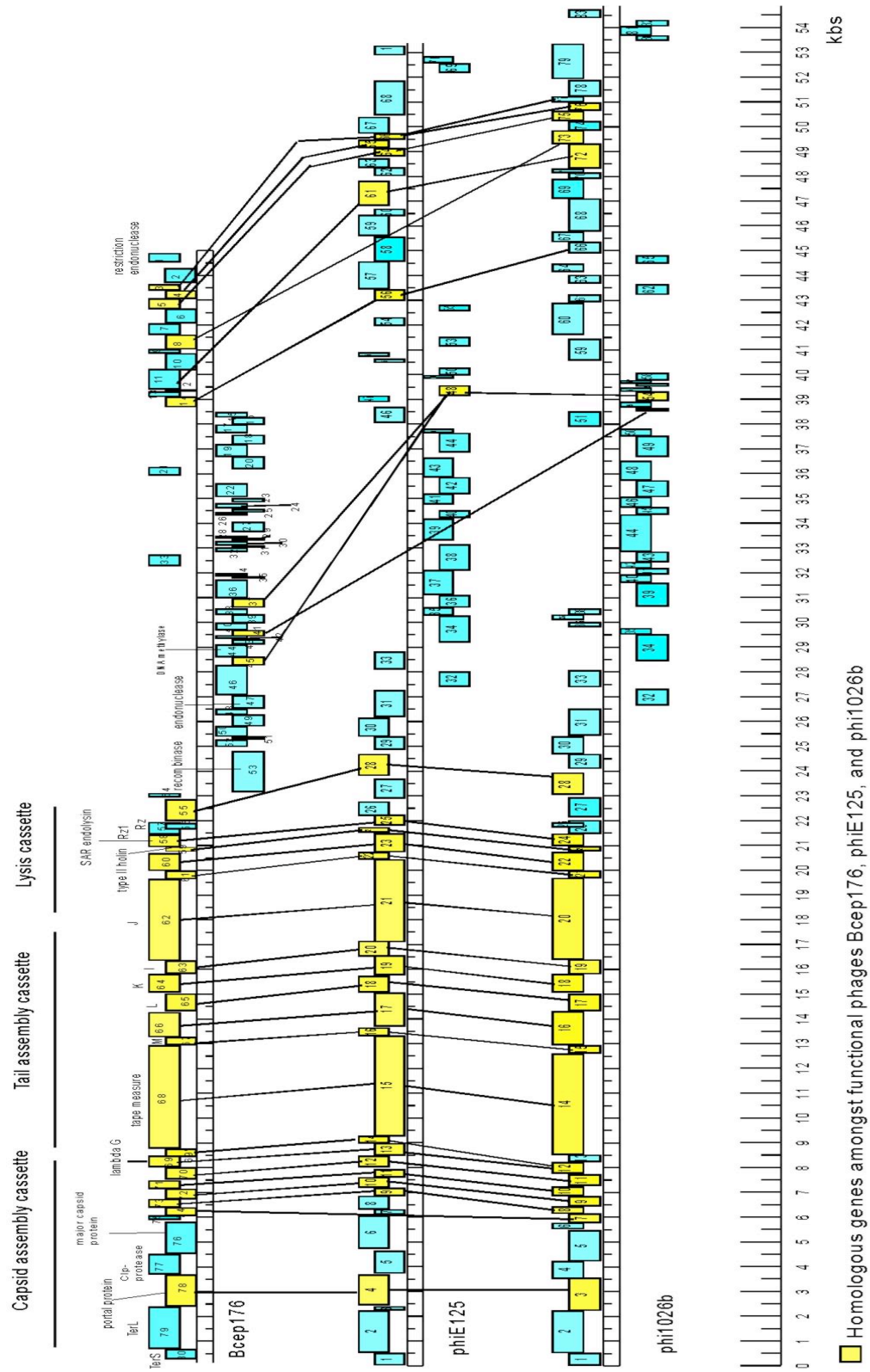

FIG. 7. Mosaicism (Bcep176, phiE125, phi1026b)

TABLE 4. Summary of Bglu421 preliminary sequencing

Contig name	Length (bps)
C	8027
B	21545
D	7483
A	5639
E	8123
Total bps sequenced	50817

TABLE 5. Summary of Bglu421 coding regions and edited proteins (preliminary)

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Cgp01	+	253	555	302	[*Burkholderia pseudomallei* Pasteur], gi|67756051|ref|ZP_00494939.1| COG1473: Metal-dependent amidase/aminoacylase/carboxypeptidase, 84, 5e-17: 54% over 50aa***	Metal-dependent amidase/aminoacylase/carboxypeptidase; Pasteur gi|67756051-like	-	MQTTATAAVSTAPISVKEVESSQIYAIGHDATTNTLAIRFRSFKTNGPTSLYHYANVPAEVFAELASAESIGTYFGQHIKPYDKKYPYVQIEKMPTIAPA*
Bglu421Cgp02	+	892	1191	299	***no hits found***	hypothetical novel protein	-	MKVTINGFIFARPARDGSLSYEFSCYDMTKWDNQAAKVRDHSIDVDVPDDFDPRPGLVENLEREKTKIRAEAEVRVTDIDRMIQTLLAIDNHPAARTPA*
Bglu421Cgp03	+	1197	1361	164	***no hits found***	hypothetical novel protein	-	MSLRPSLRRYGAHLNRQQRRAWMVQRLTRSPRVPISSGFVPRGIARVYTYCKVA*
Bglu421Cgp04	+	1363	1599	236	***no hits found***	hypothetical novel protein	-	MTTTNESSKPAGMVEVNEEQFFAVMGPQDVHPRAEPDRSVWETPRRDVLGISMPGYASSYGTPEHYFVVDRLATGAQQ*
Bglu421Cgp05	+	1596	1829	233	[*Burkholderia cepacia* phage Bcep176]gi|77864663|ref|YP_355373.1| gi|76885849|gb|ABA60039.1| gp38, 91, 8e-07: 44% over 39aa***[*Burkholderia cepacia* phage Bcep22]gi|38640345|ref|NP_944268.1| conserved phage protein gi|33860412|gb|AAQ54972.1| gp38, 70, 1e-05: 44% over 33aa***	hypothetical conserved protein; Bcep176 gp38-like and Bcep22 gp38-like	-	MSDTTIKPDDGGPAFPLSANEYAGHGPVWGMTLRDYFAARAMGGMLAADRAWSEGAHQEQLALSAYGIADAMLNARG*
Bglu421Cgp06	+	1831	1986	155	***no hits found***	hypothetical novel protein	+	MRWLDRLHEKHPVMTMAGAILVLFVVLYLAHEWDQGDSNLLRWQLAAARTA*
Bglu421Cgp07	+	2005	2910	905	*[Silicibacter* sp. TM1040], gi|69298208|ref|ZP_00620488.1| ERF, 240, 5e-14: 54% over 52aa***[*Rhodobacter sphaeroides* ATCC 17025]gi|83371437|ref|ZP_00916242.1| hypothetical protein Rsph17025DRAFT_2588, 245, 3e-24: 42% over 63aa***	ERF; Silicibacter sp. TM1040 ERF-like	-	MQTATMADVADVELNEARPEQPLTSAPAGHRTSHSTALATTTPGDLLRLAVEKGAGLDQLKQLMDLQERWEDREARKAFVSAMTAFKREPLEIFKRKLVEFRTRDGEVTSYKHAELSDITDAVGPAMAKHGLSFDWDIHQSNGLITVDCIVTHALGHSKKVTMSGAPDASGKKNAIQQQASTITYLQRYTLLAATGMSTKDQDDDDGAGGADTDGGSGGDQQQASEAQQNGGRSARAAGQNQAARPPFYDQKKFDANKDSWREVVKSGRKTPAAMVAFIESRGAPLTEEQKLTIDSWSHEND*
Bglu421Cgp08	+	2903	3775	872	[*Pseudomonas syringae* pv. Syringae B728a]gi|66046057|ref|YP_235898.1| hypothetical protein Psyr_2821, 553, 2e-68: 49% over 138aa*** [Yersinia pseudotuberculosis IP 32953] gi|51589400|emb|CAH21022.1| conserved hypothetical protein, 534, 5e-64: 48% over 133aa***	hypothetical conserved protein	-	MTERITHDLVQGSPEWEAFRLEHFGASEAAAMLGLSPKVKRNELLHMKHTGTPKEFSEWVQENILDKGHELEALARPIIELRIGEDLYPVTMSIGRISASCDGLTMNGRMAWEHKQYNEALFAAVSRGELPEEHAPQCQQVLMVTGADVLVFTVSDGTDDKCVSMEVLPDQKWFNRIVAGWGQFEKDLIAYEPRDIKEKPKAEVIMGLPTLAVQIKGEVVASNLPRFKAAAEAFIAGIKTDLKTDQDFANAEETVKTCEKAEKDLELAKSAALAQTSSIDELNAHGRRHP*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Cgp09	+	3750	4526	776	[*Pseudomonas syringae* pv. syringae], gi|66046057|ref|YP_235898.1| hypothetical protein Psyr_2821, 250, 1.00E-20:40% over 144aa***[*Yersinia pseudotuberculosis* IP], gi|51589400|emb|CAH21022.1| conserved hypothetical protein, 216, 1.00E-16:37% over 143aa***	hypothetical conserved protein	-	MRTVDAIREQLRRKRLELDKIVTKRKGEIKDEIILGGKSAYAAH VKSLNDELADVTLAVAAPDFVAAAKNKRTLASLHEAIDTAVAN GKIAADAAARDLRGKLDWYRKHAAEHTFLFRDLQQLIQKPAE DFQLAVDARIEQHKKQQAEEAERKRAAEQAAADAAAAKANA APAVDEPAPVGAQATRQPAPQTTSVPAAASAGPRRGTGGSMRA QPSRPSDDEIVGVLALHYRQHESKIVEWLLAMDLKAASERLA AAF*
Bglu421Cgp10	+	4526	5209	683	[*Bordetella pertussis* Tohama I]], gi|33571358|emb|CAE44836.1| hypothetical protein, 135, 3.00E-07:26% over 146aa***	hypothetical conserved protein	-	MPEFRFTKIKMKVTSINVRQELHGDEHCLAMDIGLEFNQSNRA LDKLDGRLVEALYWKSPTGNTEALEGVERVTDMPNLRFEHLV TPLKWAEKYEEGLFRVHHGNKPSEDIVMPESKINEIKFHPKEGG TTLFTVRVQCNPDEAQVARMCALLQSEITATIDSDPDDEDTAPP PSAAEQPEKSVAAKTGRKPRRGKQTDAFADAAQQIADGASAE*
Bglu421Cgp11	+	5303	6100	797	[*Burkholderia dolosa* AUO158]], gi|84361734|ref|ZP_00986388.1| COG4230: Delta 1-pyrroline-5-carboxylate dehydrogenase, 357, 5.00E-33:38% over 264aa***[*Azoarcus* sp. EbN1]], gi|56311484|emb|CAI06129.1| hypothetical protein, 135, 3.00E-07:55% over 63aa***	Delta 1-pyrroline-5-carboxylate dehydrogenase; *Burkholderia dolosa* AUO158 COG4320-like	-	MTMRDEIVRYLSTVLESTGKTIAARIAQTQFDVAKELNKMHGD GLVEREKRKGGNEYIYWLTGGPRRPSSPAQLPASAPVPPAPAVV AEQPPADPVVVQELLAELEAERSTSARLRGEVQSLTLELNAER VISVSQSAEMSNLKQRVAELVSLSEANMAVDVTADLSEQLIAA RSQVESAEQTIQLLRANNSALEQQIDNLTLGPVGSKGPVFFTVG RNCDSKRHTSIEKARKRANALVRSERESEVHVCTSVGCVVRG VEWVGK*
Bglu421Cgp12	+	6097	6570	473	[*Burkholderia pseudomallei* Pasteur], gi|67756052|ref|ZP_00494940.1| hypothetical protein BpseP_01001722, 368, 3.00E-34:71% over 103aa***	hypothetical conserved protein; Pasteur gi|67756052-like	-	MSQIAEDFDRATGEKTPWNPSRRAIKRVLNPLPAPTVCRFCGG KVEIVRNSQIYHGRDYGEWPWAYLCGDCRAYVGMHPFTAIPL GTLADGPTRKARKRAKAAFNPLWETGGMTRTAAYSWLAARL DIAFGECHVGWFDVAMCDRVVEVCSHPGKDS*
Bglu421Cgp13	+	6908	7102	194	[B*urkholderia dolosa* AUO158], gi|84361736|ref|ZP_00986390.1| hypothetical protein BdolA_01001853, 219, 5.00E-17:68% over 63aa***[*Burkholderia pseudomallei* 668], gi|67740330|ref|ZP_00490831.1| hypothetical protein Bpse6_01000298, 143, 3.00E-08:45% over 60aa***	hypothetical conserved protein	-	MTTTQETVWCALLRLEHADLADVDRNLLRPAFAALHGSQAIRI PERVVARIRHLDATLPKSKEV*
Bglu421Cgp14	+	7105	7515	410	***no hits found***	hypothetical novel protein	-	MLLSQIQGSDASYLASTVQTYWGGKVGTEPHPKYGYGCDAFP NGWAEIDWKEFAQSKFFQYTPKAMGWIRTTIGDAKLFFMHDQ TGFALISDYWPGTVKVFRFGCEHSMKSEMVGNCLHRYTCQTC GFSEVVDSSD*
Bglu421Bgp01	-	159	443	284	[*Yersinia pestis* CO92], gi|15980107|emb|CAC90909.1| hypothetical phage protein, 163, 2.00E-10: 43% over 88aa***[*Yersinia pestis biovar* Medievalis], gi|45436172|gb|AAS61728.1| hypothetical protein YP1489, 121, 1.00E-05: 37% over 91aa***[*Yersinia pestis* Angola], gi|77633676|ref|ZP_00795797.1| hypothetical protein YpesA_01002956, 118, 2.00E-05: 36% over 91aa***	hypothetical conserved protein	-	MFTVRVVLHSIQNKDHATYTKLHDEMEKDGFSREITGSKTYR MPPAEYSFEGTLTVVEVRDKARAVANRVHADNGVLVTQSAPD GRAWVGLKPA*
Bglu421Bgp02	+	585	1106	521	***no significant hits***	hypothetical novel protein	-	MRQIDLATLEQIELAYKAGTKSVNAIAKTFGVPEATLRRLAKK NGWVRGAPDIKRQIVADHFAGVTKGLANDEVRQSQESAAGAD IEDMERGLRLHRMCLIQLEATVAGATDPKVIKVITEAAALAITGI RKIRGLDTPTPADATDLDAAIEAELAHMERRRQAGAAPDAEGS *

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Bgp03	+	1168	1584	416	*[Bordetella* phage], gi|45580769|ref|NP_996635.1| hypothetical protein predicted by GeneMark, 191, 9.00E-14: 45% over 98aa***	hypothetical conserved protein	-	VILYGGAAGGGKTDLALGKALTQHQRALILRREYPQLMGMID RSREIYSEYRTFNQKGVWRCEFTGKKRLINFGSVQYEDDKQKN QGRQQDRRVFDEEANSSSRRWSSSPAGTGARIPTSSARCCSART RQPMRPATG*
Bglu421Bgp04	+	1614	2570	956	[*Bordetella* phage], gi|45580769|ref|NP_996635.1| hypothetical protein predicted by GeneMark, 634, 4.00E-65: 48% over 330aa***	hypothetical conserved protein	-	MHPNPAAPGELRYFASVAGQQIECADGRPFIIVEQENGDQVLVY DFDATAIDAVDIIRPLTRTFIPARVTDNPFYAQSGYVAKLQALPE LRSKMLYGDFAAGREDDPWQVIPSEWVRLAQERWRARSRPRI PMTALGVDVARGGQDQSIYTPRYGNWFDEQVCQPGLATPDGF VVAQQVFNLREPSTLVNLDVVGVGASPFDIIHQVIGDKIWGISG AARTDELDMSGQFGFVNLRALLWWRMREALDPINGEDLALPP DPALAADLCAPRYRKAPRGILVESKEEIKKRIGRSPDRGDSAVY ALPERSGGFAFGSL*
Bglu421Bgp05	+	2632	3939	1307	[*Burkholderia pseudomallei* Pasteur], gi|67756042|ref|ZP_00494930.1| COG5518: Bacteriophage capsid portal protein, 1096, 1.00E-118: 49% over 423aa***[Rhodobacter sphaeroides ATCC 17025], gi|83368376|ref|ZP_00913240.1| Phage portal protein, HK97, 159, 4.00E-10: 28% over 389aa***	capsid portal protein; Pasteur gi|67756042-like	-	MKRAVAGAVRAIRAGRVGRPKHEPQRRTSAFARLRQTGSWGL RDARPLIKPVPANLRYFSKTPYASRAIQIYTRSITSLEWAVKVKK DVAENSEIKRQLEVTSACLFSPNNQDTFTTLLQAVIEDLCVCGA GAIEPEVGGDKLRPLWMWPVDALSIQIYADWDGDEAKPRYCQ TYGYGNVGVAQGVNLLNRELVYIRDRVTTDSPFSFGMLEIAFE SINRLLGVAEYAGDIAANAHPENLIFLKGASQETIEAFRTYWRN EIEGQGQTPIIGGDDTKVENLRGSTDEVLFLKYQEFVIREIAVAF GISAQNLGVEQNINRNNGEVAEDRDWDLSIKPMARAIASYLNR EVIWGLLGHTQIELTPGGLDREDEKATADIYTLEYKGNAITPNE YRARRNMPPLKSKFGDMVFADAQIAIEAARGAKQVNPELTNIE *
Bglu421Bgp06	+	4005	4313	308	[*Burkholderia pseudomallei* Pasteur], gi|67756038|ref|ZP_00494926.1| hypothetical protein BpseP_01001707, 168, 4.00E-11: 66% over 50aa***	hypothetical conserved protein; Pasteur gi|67756038-like	-	MGLPIHTVEVLPGSNPPAQPDRVAVALLTAVATAAGSGAGAAV TTVTGLSLPPTYNVQVTPNQDAVAYVTGKTQTGFSVVLNPRLA ANTLAAGTVDVAIFA*
Bglu421Bgp07	+	4371	4580	209	***no hits found***	hypothetical novel protein	-	MGVRTSILAITPLTAETSTAAANGQDLPGMLNFAAEKVADAIQV LNAIADFLPAGSNLTAIQAQITALT*
Bglu421Bgp08	+	4631	4885	254	***no hits found***	hypothetical novel protein	-	MAKVNSDAAKGAAAEAKQPAKEPSLYQALEKVAEKALTDGD HAAYQVLHPVVVALAGAKFAAGQAARTPAAQDALDLLERVKS L*
Bglu421Bgp09	+	4894	6357	1463	[*Burkholderia pseudomallei* Pasteur], gi|67756035|ref|ZP_00494923.1| COG0840: Methyl-accepting chemotaxis protein, 833, 3.00E-88: 50% over 456aa***	methyl-accepting chemotaxis protein; Pasteur gi|67756035 -like	-	MMSLSKEERDKLPADQFAVPGKRKLPIHDAPHTRLAWDMVDR TQGLSDTERREARDNIRRRAKELGIDTTGWHAVRAAALAAAT AMTLWGMSLNVPEVTDHPNRMPFSGVLTFVNQPSDLPPGGSG GKRTYLPKDVAEKALGSLLGMAVDFSDDLSAHDVTQKIGVITG AEIVGDEVRIEGFFYAADFPQECARIQEEKEDLGFSYEVRAQTR AMGDLLQIVSCVFTGAAVLYKDKAAYQSTSLAAQAEQDIEMT KEEMEALLAGALGPITKQITDVTAQVAALKASTETAVQANAA MRDRVSPHATALRNCAAGMEAAGIGLHSARGHVKVLHHMAA SMEAEAAGGHLPHIYRDHDWGFSASSDPAELAAAAAAAAAR AKPADASGLTADNPAIKAITDTLASLGTQLTDIKAAAFKSAEAP QRRTISSETIALLAKGGLTLTDEQKKEGITEEQIDATLKAAGVTN ITARIAAKQKIAADGLLRK*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Bgp10	+	6407	7465	1058	[Burkholderia pseudomallei Pasteur], gi|67756034|ref|ZP_00494922.1| hypothetical protein BpseP_01001703, 991, 1.00E-106: 63% over 310aa***	hypothetical conserved protein; Pasteur gi|67756034-like	-	MSKLIARSADTSTDDALLALGNVLGASGNTVQALSAAADYSG PGALEVPVFEREIVDLIRRESPMLAITEHNPATGHPHRYFEQIAIA TASSNDPRNLAATPTGPTRVERAAFIKATVAQSNLSLFDRDVTE QQGQFATLQAKDVEDILTAIIVWRANMYWAGTDTSLLMPTTLQ WVGGLEQITQQATIPYGSSIIDGLKTMVATMMANKIFKPKPTAI GLNPLLIDKIEKEAKASHIELRTKEIVAGVTVKYLATQAGELPLI PDPYMPTDTTGKYGFANPPNGLSNYYAAILTMPMVEVVYIGK GTNGQPRIFQLGLTGNLAGQFVAVQFDALLFKGYSYAHAVVAV VSN*
Bglu421Bgp11	+	7547	7831	284	[Burkholderia pseudomallei Pasteur], gi|67756033|ref|ZP_00494921.1| hypothetical protein BpseP_01001702, 180, 2.00E-12: 48% over 75aa***	hypothetical conserved protein; Pasteur gi|67756033-like	-	MNLYLTTGKRDHLVICQPGMEFPTSDFLAEDGHPITFKVQFIDG KAKDVPTNVGNYLIDKGVAQRSPIMTDLAMAKRLEVHDERVR THIMLASQ*
Bglu421Bgp12	+	7828	8220	392	***no hits found***	hypothetical novel protein	-	MTQRTDISGVGGELGPALQCLQSAAQLGLTGTVPANGSLAGPA LVSNGWKLFSVGLKSSQAGQISVQRFLDLAGTVPQGAAVTASL TANTAASATVGTDGLPFASVQVTVTNSSGSAATLSAVQGLFQA S*
Bglu421Bgp13	+	8222	9160	938	[Burkholderia pseudomallei Pasteur], gi|67756031|ref|ZP_00494919.1| hypothetical protein BpseP_01001700, 614, 8.00E-63: 46% over 308aa***	hypothetical conserved protein; Pasteur gi|67756031-like	-	MPSSYLSDADYATYGLPASTTANQVTQASTIIDLYLKRPDGLVW MPDGTGAPGWMAAKSPTITYQAAGPIPSGQNVQVEITGGVAA VQVGDVVILDRTNSQLAEPVAIVSFGDGLQSIAGVRPPVTVTLQ NVVFDHPGPCTLDAGMVVKQHKYLPEGRPLTNLAFTPIARLLA GQGRYAYGRRGASSRYQVDEFNLIASLSHFGGPPVWEFFPMVN TGVDFDTGQVWVPAGVMLAYYSEVNLWYVAGFPASGLPPAVF MACANIIQAKANSMQLGATKKYQAGDTSLERFAATLIDDDTK AMLAPYMAKAWS*
Bglu421Bgp14	+	9157	9510	353	[Burkholderia pseudomallei Pasteur], gi|67756030|ref|ZP_00494918.1| hypothetical protein BpseP_01001699, 163, 2.00E-10: 35% over 113aa***	hypothetical conserved protein; Pasteur gi|67756030-like	-	MSFMYPRTISISRQAVSTAGGLKPYSGADPVAEQVLFSGLPASIQ QTSTSAKPDANLPADARSRSIWRVFMPLSAGVPAGAVLRGDIV TDDAGQRYQVWAPYVNSLGPNLLVERLEA*
Bglu421Bgp15	+	9511	10353	842	[Burkholderia pseudomallei Pasteur], gi|67756028|ref|ZP_00494916.1| hypothetical protein BpseP_01001697, 454, 3.00E-44: 41% over 249aa***	hypothetical conserved protein; Pasteur gi|67756028-like	+	MADVQDVLDTVAAQVAGFVYPNGTGQPSVSGKAINVFPGWP MPNQIDADMPTGVADVSVYATPTERNTTRYQPKQKVMSIAVA LTLTAAGGTLTVGGVMPSPFTPHNVAALINGQAFIYPVQPTDSL TSVATGLANLIAAQYPGTTNSGPVITLGAGVEAAVARVGTTGV VTTEWERQAQLFQITVWAPDPATRKAIGAAIKEALAPITFITMP DGYGARIRYVRNVLSDVAEKVHVYRHDLYYEIEYATTATEQVA TVVATKVEVETQAGTPVFTRTY*
Bglu421Bgp16	+	10362	10541	179	***no significant hits***	hypothetical novel protein	-	MATKDPADDTAAQTAADRTLVVVHAFGDYRRGDSITDPSEVA VLASENANHCRKVIPQ*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Bgp17	+	10603	12603	2000	[Burkholderia pseudomallei Pasteur], gi|67756026|ref|ZP_00494914.1| hypothetical protein BpseP_01001695, 276, 1.00E-23: 37% over 198aa***[Acidiphilium cryptum JF-5], gi|88939387|ref|ZP_01144836.1| phage tail sheath protein, 292, 2.00E-25: 42% over 166aa***[Gluconobacter oxydans 621H], gi|58002085|gb|AAW60979.1| Phage tail sheath protein, 179, 2.00E-12: 40% over 140aa***[Gluconobacter oxydans 621H], gi|58003285|gb|AAW62179.1| Putative phage-related protein, 179, 2.00E-12: 35% over 139aa***	phage tail sheath protein; Pasteur gi|67756026-like	-	MGTMIYQYGQLNTAGAMAPGAYVQIVKPPPVVAGVATNGYGI VGVASWGPVNSPVVTGSPQANQSNWGAVTNRQRDLATAIAIA FAIGQYNNVAVRVTDGTDTAATGTLKDASNAAGATLTGYYTG LGNSLVATPSTGTKPNSYKVTLALPGFQPEIFDNITQGVSGATVT PGTGYTSVPNLAISAPQAAGGVQAAANASLKVVSANVSGGGA NAGSGYLTGDTVTLVNGVVLGVTASAGAITALSVTNGGSLTGG SVPTNPVAPQSTSGGGTGALINLVWGLGPVSITNPGSGYTSATA TPSGGGAGTGGAVTLATSIWLNLVNAINQGQSGVRGPSKLVVA TIGTSTDAPALTAVTLSGGTDGAAGVTDATLVGTNTTPPTGMY. LQSAGVQTMNLVDHSTSSQWSTMAAFAQQFGVFGAAQGQPG QSIATVSTNLNTAGVDTYGLKCLVGDWVYWQDTANNVQRLV GPATIWGPMRANLAPNQSTLNKPVLGIVGTQRSVQEVPYSGPE TLSAVQARLDFLANPAPGGNYFAFQTDRNVSSDAATNSEAYTT MTNFLALTLAANYGFVVGNPQTTDLRTEVRDSITSFLTNLWLV SQYIGDVNNPQAVPFKVTLDASNNPDAQVAIGLMQCLVVVKY QSIVREFLISLQGGSTVQVTVSNGS*
Bglu421Bgp18	+	12718	13194	476	[Escherichia coli CFT073], gi|26107709|gb|AAN79908.1| Hypothetical protein c_1439, 217, 8.00E-17: 44% over 106aa***	hypothetical conserved protein	-	VSLDGDRPVTDSRKVAKQFGKQHGKVLRDIRVIVKQCGDEFA QANFGFSHEINMLANGRRDPVCSMTKDGFMLLAMGFGGAKA MAVKVAFIKAFNAMAEALRTGLWQRRREAQAAFDAGRAQASI DGSGLSRWRHQKHQHERLIDELDRQLDLPLQLT*
Bglu421Bgp19	+	13302	13814	512	[Xylella fastidiosa Temecula1], gi|28198900|ref|NP_779214.1| hypothetical protein PD1002, 258, 1.00E-21: 49% over 103aa***[Escherichia coli E110019], gi|75235791|ref|ZP_00719946.1| COG3617: Prophage antirepressor, 210, 6E-20: 47% over 56aa***	prophage anti-repressor	-	MMKELMFQNQSIRLIEQDGKQWASAADIARALGYAQADKVTR IYDRHKAEFSASMTMLTMIETAAPQSGGLLNKTTARVFSLRGA HLIAMFARTEKAQAFRRWVLDIIESHQASQSLIQEWFEAKAAV DAQDKFASLCGKGLNEHKRRKPPLAERLGTIAEKIQPSLSFN*
Bglu421Bgp20	+	13854	14294	440	[Burkholderia pseudomallei Pasteur], gi|67756025|ref|ZP_00494913.1| hypothetical protein BpseP_01001694, 224, 1.00E-17: 35% over 143aa***[Acidiphilium cryptum JF-5], gi|88939386|ref|ZP_01144835.1| putative phage protein, 140, 7.00E-08: 30% over 143aa***	hypothetical conserved phage protein; Pasteur gi|67756025-like	-	MAGINSFNVGRDSAQLTIIDSNFGTVTINGITKFDAKPAVVKLK SVGIGGRILHRTVPDGHALSFELDRQDPSYEQYFANAEASYFA GLPPAAIFLTQTINNLDGSISQYQYLDMALAPEDDGSWEGQAK VPQKFSAEAGRKIRLA*
Bglu421Bgp21	+	14304	14714	410	[Burkholderia pseudomallei Pasteur], gi|67756024|ref|ZP_00494912.1| COG0143: Methionyl-tRNA synthetase, 248, 2.00E-20: 43% over 126aa***	Methionyl-tRNA synthetase; Pasteur gi|67756024-like	-	MDQKVIVRPQGAPGEAAAVPMKPSQTVISQGASEVTVVDTAG RAIVLRKPGPLARLRFIDAMGESSANKLWAGTVWPLMFVVSID GVAVTPPAQKAQIEALYQRLEEHGYEALSEGVSAHFSADQSNV NEDLAKKS*
Bglu421Bgp22	+	14881	15363	482	[Burkholderia pseudomallei Pasteur], gi|67756023|ref|ZP_00494911.1| hypothetical protein BpseP_01001691, 282, 2.00E-24: 47% over 138aa***[Ralstonia solanacearum], gi|17429500|emb|CAD16186.1| PROBABLE BACTERIOPHAGE-RELATED PROTEIN, 271, 5.00E-23: 46% over 132aa***	hypothetical conserved phage protein; Pasteur gi|67756023-like	-	MRQFKSFGAFAAHLERLALAGPEVIHLIADRAAEDIEKNAKGM LGEYQSAVGQFPAWEELAESTQAERARLGYSENDPGFRSGEAQ ASIGRHVEHGHAVVGSSDQNLVWFDLGTSKQPPRPVFGPAAIH ATPRIMAMAAWTTFAWVSGRGWRRGRLTSGG*
Bglu421Bgp23	-	15347	15598	251	***no hits found***	hypothetical novel protein	+	MRIAYRIHEWLADHISWVQYPKPRLRSLSGHPHGFRAWWAHR PPINRLVAIFGPTALLFVPRIGVYLTICSILFLFAYFNRRW*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Bgp24	+	15633	17543	1910	[Burkholderia pseudomallei Pasteur], gi|67756022|ref|ZP_00494910.1| COG5283: Phage-related tail protein, 391, 5.00E-37: 31% over 365aa***[Burkholderia pseudomallei S13], gi|67760371|ref|ZP_00499093.1| COG5283: Phage-related tail protein, 360, 2.00E-33: 28% over 346aa***[Burkholderia thailandensis E264], gi|83717435|ref|YP_439523.1| probable phage-related tail transmembrane protein, 359, 3.00E-33: 28% over 346aa***[Burkholderia pseudomallei K96243], gi|53722115|ref|YP_111100.1| bacteriophage membrane protein, 357, 5.00E-33: 28% over 346aa***	tail protein; Pasteur gi|67756022 -like	-	MINVYGIGTTLKLVDRVSPELLKLSREFSKVDAIALQLGKRLQK LGAEAGGVRNVAAASKALDTSLKAVGGHAAIAERNLRGIKGA IPTGKIGIEAELTAANVQAKTLATTLAGMRGGRWLPSSGGGGG GRGGNSSHGGRIHGGNLHFGSHGVGIGGVGMGLMSDALVPLG AAMVAGYVGKQFYESAKDLQTEHARFKLFGMDDAQNKSAFD FAHNLHAYATSENEAMHYMNEAQAVFRESGETGEKALAGAK MAAPFLAKMMAASSILSEESQAKLKHESLAMLRAVELQGGAR DSQRFAELADFGFRMTQTSGGQVNWEQLRQMYRTGGVAVQR MGMDAIAELEPIIGEFKGGSFATAMRTAYNRMNGIIKLPNQAVH ELIKAGIWDKSKVELNGNGGIKRFLGNPLQHSDEYANNAAEY YFKYVRPYYDKQGYKDTDRDRMNAMFFGSTGGNLFSKIDQQ EPVIRSALEAFKKTFGLDKALDISKDTATGAEKDFAAAWTDFK TEFGNTMLPAITSMLKTGTAILRMINSAPPVFETAAKAQMDQV PGADHETVGRWGKLGDLFGWNGAKGDSANPVSGKRPVTVNV TANTILDGQKVASVVTKHQANMLGSGLYALGVDPTVAVPMPG KGG*
Bglu421Bgp25	+	17545	18510	965	[Burkholderia pseudomallei Pasteur], gi|67756021|ref|ZP_00494909.1| hypothetical protein BpseP_01001689, 375, 4.00E-35: 41% over 238aa***[Gluconobacter oxydans 621H], gi|58002089|gb|AAW60983.1| Putative phage protein, 156, 1.00E-09: 32% over 216aa***	hypothetical conserved phage protein; Pasteur gi|67756021-like	-	MATTLILGDFEFQDFEVPEVIGFGGDQRLSVKKMLGGVRDVQ ALGTDPRPIEWAGQFFPTQDGQTALDRALTLKQMMDAAQPLY LSWDELYLQVFIRSFDPDYRFARIPYRISCEVVADLTAPVYEAGE PDADDLIDGDLDSANSLTEGIGDSTLSGLMDTVSSAVGSVSTFV GAAQSTVASVLQPLNAAAQRVSSLIASTDSVLASVGVPAGVLP SVPLLSNISVFTSQIGACGQQTQLLQLSGLLGRMKTNLGLVNSS VRTTTVPGGNLFDIASKEYGDPTAWTLIANANNLTDPTLAGIST LIIPPYVASASGGILSS*
Bglu421Bgp26	+	18514	19632	1118	[Burkholderia pseudomallei Pasteur], gi|67756020|ref|ZP_00494908.1| hypothetical protein BpseP_01001688, 816, 3.00E-86: 47% over 367aa***	hypothetical conserved protein; Pasteur gi|67756020-like	-	MTSFLQPTARSPRGAVKLNGELIKGWIELEVDNNAYYAADTFR CTFAGAQLPPDRNVAWFSEQQDMYVELFIGFPADPLNFSPSDL KSFIYGQVDHVVIDPAAYTVSVDGRDLSRVFIDNKTTQKWPNK TSSEIATALAKSHGLTPVVTATTTQVGKYYEIDHVNMADERSE WDILNYLANIEGFKVWVRGQSLYFQAPLDPTKTTPYPIVYVPP GTTAGAKANFEAARFTRALTISRGIQVKIRSWNKSYAKGFTVAY PSNVKTIRVGSSAVGAGAQIYSRTVPNLTQDQALQYAQKWYQ QIVAHEMKLEGVALPGDNDLDIPSMLSLSGTGTKFDQAYYTDN IRRTLSFEGGYDMIIGAKNHSPDSEVTL*
Bglu421Bgp27	+	19629	20348	719	[Acidiphilium cryptum JF-5], gi|88939381|ref|ZP_01144830.1| Phage baseplate assembly protein V, 236, 5.00E-19: 35% over 160aa***[Chromohalobacter salexigens], gi|67676656|ref|ZP_00473403.1| Phage baseplate assembly protein V, 193, 5.00E-14: 29% over 234aa***	baseplate assembly protein V; PasteurH-like (see *Acidiphilum* gi|88939381)	-	MSYAALANAVRQQALGAQGQQSIPRLATISSYDPNRHVVKVIV QPVDPEVGEQESNWMPLGAIGIGNGWGVAVGPQLGDQVLVEY EGGDFSSGTVVSRIFSIAQPAPPVPSGEIWAVHEKGQFLKLTNDG KLTANDAAGSTVVMNGDGTGTVTFAGGLKVNADTIINGTLTV TETITGEAGITITGENPTGAASTITGDFKIVGNIDSTGSITNNGKH VDSTHAHSNGNGGENTGTPVR*
Bglu421Bgp28	+	20345	20506	161	***no hits found***	hypothetical novel protein	-	MMVDLYHYYGTDLSESPSADLAVPEDQTTGEQRVHRSLLTNPA LSDTTRNPIA*
Bglu421Bgp29	+	20579	20755	176	***no hits found***	hypothetical novel protein	-	VTEALIKGQMLQKAAVARQPHSQISMTQTQKVVSTSISYTDAN TAQTQLVSFDISQEP*
Bglu421Dgp01	+	380	1126	746	[Xylella fastidiosa 9a5c], gi|9107484|gb|AAF85116.1| hypothetical protein XF_2317 , 422, 4.00E-40: 42% over 229aa***[Xylella fastidiosa Temecula1], gi|28199236|ref|NP_779550.1| hypothetical protein PD1352 , 378, 6.00E-35: 46% over 175aa***	hypothetical conserved protein; PasteurE-like (see *Xyllela* gi|9107484)	+	MATGDSTDILSRLKSYMPRGWFGDWAEAPIINAVLSGVAAVLS TIYLLVQFAWAQTRLQTSSGGWIDLWAADFFGDSLLRKAGETD ASYIARIKLTLFQRRATRPAMDAVLTQLTGRHPVIFEPARPLDTG CCGANTGPNSFCGVARMGSIAAPFTALITVYRPQVTGGSLGAA YANAPTWSALSAPTSHGYTGSLTQEVTQATDADIFAAVNATRP AATKIGVCITNFGVPTPVLGANFVLDQSTLD*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Dgp02	+	1196	2836	1640	[Xylella fastidiosa 9a5c], gi|9107487|gb|AAF85117.1| hypothetical protein XF_2318 , 552, 4.00E-55: 46% over 267aa***	hypothetical conserved protein; PasteurD-like (see *Xyllela* gi|9107487)	-	MRRIETYIGQQVYEWTFSAQAQQTMVGLAKLSSAVLGTNTVV LGLACAPTSPAGMTVQIGKGEIYQLENLEATACGTLPADTTDQI LKQGIQLGTYTSPAFAPPGTSGQSISYLIEAQYQDSDISLDPTTG NAPVVLQFYNAANPATPWSGPGNSGATSNTFRDGVVAFHVKA GIPATSGTQATPSPDTGWVGLWVITVAFGQTSITSANIAQYSGAI ILPSGLLQSVQTGNLNFGTDVGAVNALQANFPVPSTTLTDNQA FWVKVANTNTGAATFTPNPGVIAAAPIVGVSQSALTGGELAAG GRALLLWRADISSYVLITCTGSVQLGAPRFQIFTSSGNFTVPPG VTTIYRTGCAGGGGGAGSPNIPANNIVAGSAGGGAGRSWIRNP MSVTPGQVVPIVIGAGGVAGAVGGAGGNGSNTSVGSDVLPAG AGGQVGSSGTPTQAWPGQPGGNGFPNGEYGQDTSQYGPGATG GRGGSGPFGTSGAPGRGAIGGVANLIAPSASYGFGAGGSGAGG SYGPTTVTGTTNGSVGTPGMPGFMLIEW*
Bglu421Dgp03	+	2830	3057	227	no significant hits	hypothetical novel protein	-	MVITMARYAYFDAQVKAVLGWIDTEAYGYAEDPLKAFPESQM LPITDDADWHPNDGQRWYVVNGSLTNVAPTSSN*
Bglu421Dgp04	+	3073	4677	1604	[Aspergillus fumigatus Af293], gi|70988641|ref|XP_749179.1| exo-beta-1, 3-glucanase , 121, 4.00E-05: 23% over 331aa***[Burkholderia fungorum LB400], gi|48784624|ref|ZP_00280990.1| COG5295: Autotransporter adhesin , 121, 4.00E-05: 22% over 409aa***[Burkholderia vietnamiensis G4], gi|67548816|ref|ZP_00426697.1| Parallel beta-helix repeat , 121, 4.00E-05: 24% over 209aa***[Mesorhizobium loti MAFF303099], gi|14023267|dbj|BAB49874.1| mll2848 , 119, 6.00E-05: 21% over 346aa***	exo-beta-1, 3-glucanase	-	MNRIFKTLLLAVSLIVVAVPRLAMGQIVQFVPGQVLTAQQLNTA FSRIVASTLPVAGGTLTGPLNGTTASFATGNFSTGNFTTLTSTGN VSFSTPLSFASGGTGANTALGATSNLQFQASLSGAGGRSVASKL SDVVSILDFPGCDKTGVADSTACIQAALNSGAKTVYIPAGQYRE SGLTLPQTQGFTLYGDGPNSVLIQTGGSISYPAISGAFTFDSHSTI RDLKFDGTAGTANTLDTTFAQTLDLLNLAFNNVPVGYSSIKID GNPTTGVYMHDVRLKNIRIYSTTAGKAGIELGAYASDSTIDGFI MNGGFVVNYCLYANPGAQTTMVQNSHPYNAAINVVRLAGNN NDFGFVGDTIDNALGDVFYIKNSTHTRISSTWIESINSFQRGLVL DGSSNNTVMGLGCQTYGVTNATSCVAEINGAAGNQIFGAQPD SSSNYSTPFSLSGAGSFYQSVNSGNSLNAMSAASGTSPSTTATG SDGSIPITLNPKGNGGVVLNFSNLIIAAFYSDTDGELALVVYNT RHRLRNNT*
Bglu421Dgp05	+	5166	5489	323	[Burkholderia fungorum LB400], gi|48782904|ref|ZP_00279384.1| hypothetical protein Bcep02005878 , 138, 4.00E-07: 33% over 100aa***	hypothetical conserved protein	+	MNQMEAQVRSHLERQDAHLRQQDEHLRIQDGKLDTVVANTQ SMVETWQGGQKVMRALCRMADAWRFLVRNVALPALGFGAA GVILLRYVQHEPIPEWAHAVLKMLIG*
Bglu421Dgp06	+	5486	6061	575	[Burkholderia fungorum LB400], gi|48782903|ref|ZP_00279383.1| COG3179: Predicted chitinase , 695, 1.00E-71: 70% over 187aa***[Azoarcus sp. EbN1], gi|56315596|emb|CAI10241.1| putative lysozyme , 383, 1.00E-35: 45% over 188aa***	lysozyme	-	MNAMTPQILSTALQIPLARATPFAAPIADAMARYAVDTPARQA AFLAQCGHESGRLQWLRELWGPTPAQRAYEPPAAKARELGNT QAGDGFRYRGGGLLQITGRYNFRVYGQKIGVPLEASPDLIGQP TTAALVSAQFWADHGLNALADAGDFLGISRAINLGNARAAGT PNGMPDRLALWASCKAALGAA*
Bglu421Dgp07	+	6147	6671	524	[Burkholderia pseudomallei Pasteur], gi|67756015|ref|ZP_00494903.1| COG5256: Translation elongation factor EF-1alpha (GTPase) , 144, 8.00E-08: 31% over 185aa***	hypothetical conserved protein; Pasteur gi|67756015-like	-	MHPINGAASDAAAAHVTAAHAERETLAVLVNIPEHAERTMTAI FASTRKVLIEREGGRCFICNATAQEAGHPLEAHHHPIERSFAELI DWDRFKAAALAGEFGERVRAFDWAAFTHWEQFVDDMTVNG LLLCKAHHIGKDEGIHALPFPIWIAQRYAREGYQFSDIEIIHHNQ E*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Dgp08	+	6675	7031	356	[Burkholderia vietnamiensis G4], gi|67548969|ref|ZP_00426846.1| hypothetical protein Bcep1808DRAFT_0850 , 177, 1.00E-11: 65% over 66aa***	hypothetical conserved protein	+	MPMKSSLVTGGITIGVTDLVPTVDWVLGGCRGPVPASLSSLIAG LLAAAFHAGVNWFAERAESRSADGASGAIVVPKLGTMVTVV GAGAASGGGGGMPSPAAPSGSVAPTPASPDTAT*
Bglu421Agp01	-	906	2042	1136	[Bacteriophage phi1026b]gi|38707923|ref|NP_945064.1| gp33, 781, 2.00E-81: 44% over 366aa***[Bacteriophage phiE125]gi|17975195|ref|NP_536390.1| putative site-specific integrase, 767, 9.00E-80: 42% over 361aa***[Pseudomonas putida F1]gi|82737165|ref|ZP_00900017.1| site-specific recombinase, phage integrase family, 280, 3.00E-23: 27% over 361aa***	site-specific int; phi1026b gp33/phiE125 gp34-like	-	MTPRRRKSPDGLPSRVYIRDGSFYWVRPDNNKWIRLCRVSEGD ALMLERLAAEKRQFTAAAGGGNLPRLVGEFMDTHKMRYAPSF RDEWVRRGEAVIESLCEWNVERVDAGAVEDFLMKNWADRLP MQRAMKAWLSTFFNWIIRRKKLMQVNPCREVIVKKPKARDV YIPDRDFVAIRDALMTYTYTKNAGTPKEKTLVGKVQTGPMMQ CFIDLCYLTCQRSTDIRLLRWDQIDREAGVIHFVPTKTEDSSGEA VDWPLTPDIEAVLARAKLLDPAFGQTYVIRDRKGKPKTDQSCR DAWDGAIDRVAKLDPSIAGKPYTIKDIRAKALTDAKRAGYAIE DLRVAGAHTDTATTETYLKAREVPRSKVVLHLPMPKSA*
Bglu421Agp02	-	2039	2287	248	[Bacteriophage phiE125]gi|17975196|ref|NP_536391.1| hypothetical protein phiE125p35, 159, 3.00E-09: 46% over 60aa***[Bacteriophage phi1026b]gi|38707924|ref|NP_945065.1| gp34, 155, 8.00E-09: 45% over 60aa***[Burkholderia dolosa AUO158]gi|84361747|ref|ZP_00986401.1| hypothetical protein BdolA_01001864, 104, 0.007: 41% over 48aa***	hypothetical conserved protein; phiE125gp35/phi1026bg p34-like	-	MSTGLMTPEDLVKVTHKKRYSAQAKWFKAQYGVDVPQRTDG SIVLTWETLDALAEAKLGIRAVDRRIDASPDRPPIHVLRKA*
Bglu421Agp03	-	2284	2922	638	[Burkholderia dolosa AUO158]gi|84361746|ref|ZP_00986400.1| hypothetical protein BdolA_01001863, 542, 1.00E-53: 61% over 157aa***[Bacteriophage Mx8]gi|15320584|ref|NP_203428.1| hypothetical protein Mx8p14, 261, 4.00E-21: 33% over 231aa***	hypothetical conserved phage protein	-	MRDLPIMFTPDMAAGTRSGVKTQTRRPMRPQPYDDHGRLAW MFAKNRGSWGPPGGAPDPAWLARCPYGGPGDQLWVRETFVA FGRWETRFNVKKGRDEWHFVDMTLDTGRQYRYGGDLPNAQR AGATPTWWRRPAIFMPRAASRTQLEVVRTRIERVQQISEADAIA EGMSLGDINPVGSYAELWDSLNAERGLSWAANPWVWVVEFR RMEG*
Bglu421Agp04	+	2992	3237	245	***no hits found***	hypothetical novel protein	-	MKAEYVESGYRVELNASFDADLAEGDRVVVRVAVLDAQEAAV WTKIQYVAKLPGSAPIDLETGIAAGRREAKRAIESGFKE*
Bglu421Agp05	-	3278	4282	1004	[Burkholderia dolosa AUO158] gi|84361744|ref|ZP_00986398.1| COG4422: Bacteriophage protein gp37, 1114, 5.00E-120: 63% over 330aa***[Bacteriophage phiE125] gi|17975198|ref|NP_536393.1| conserved hypothetical protein gp37, 1107, 3.00E-119: 62% over 348aa***[Bacteriophage phi1026b] gi|38707928|ref|NP_945069.1| gp38, 984, 6.00E-105: 58% over 353aa***	hypothetical conserved phage protein; phiE125gp37/phi1026bg p38-like	-	MSENSKIEWTDHTFNPWEGCQKVGPGCDHCYAEARNARFAG GTPINWGPGAPRRHTSPANWRKPIAWNAAHAEFYSKHGRRQR VFCASLADVFDNAVPASWRVELFYLIAKTPNLDWLLLTKRIGN AEEMIRESVQYVANTDEPWPWANVWLGATVVNQDEADRDVP KLLATPARVRFLSIEPMLGPIDLNQALPPFHCDTCDADYGALSA HASGTPIGIDWVIAGGESGPGARPAHPDWFRSLRDQCATAGVP FLFKQHGEWAPGSGDFGAGRFATAAVAFDGRVVPGGYRAQDY PAGASSADGWALVHRIGKRAAGRLLDGRTHDEFPENTR*
Bglu421Agp06	-	4279	4515	236	***no hits found***	hypothetical novel protein	-	MSTRTIIEINHDFLQQLIDDPLRLAVTLRSVCCDHQAELNDDNG RGRPLNLGDGIVIVYRRHHSENVRVTAKYVDIDL*
Bglu421Agp07	-	4512	4886	374	[Burkholderia sp. 383]gi|77969765|gb|ABB11144.1| hypothetical protein Bcep18194_B1030, 390, 4.00E-36: 76% over 109aa***[Burkholderia cepacia phage Bcep22]gi|38640354|ref|NP_944277.1| hypothetical protein Bcep22p47, 282, 1.00E-23: 60% over 109aa***	hypothetical conserved phage protein (Bcep22 gp47-like)	-	MTVYVDDMYRHQVGRFGNMKMSHLIADTTDELLAMVDEIGV NRRWIQHPGTRDEHFDIALVKRAAAIAAGAVAITFQECGAMNK RRRVTGQLGSPEDAVEWMKAFVEQRGREKRARAAAEAAKS*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Agp08	-	4883	5293	410	***no hits found***	hypothetical novel protein	-	MTETEQLAEARQLIAHASEYAISALDDDWHERASALLALPAPAI PESDPPLVYPNDLTPELLDVLGRPNFWCGPIAHAMRDAGADIR RKSEDEQAHVLHWLVKLVLGHGAAWQAAAGAALSQICEKAV AARKGEKS*
Bglu421Egp01	-	137	712	575	[Bacteriophage phiE125], gi|17975214|ref|NP_536409.1| putative repressor protein gp52, 5.00E-07: 32% over 126 aa*** [Nitrosococcus oceani ATCC 19707], gi|76882455|gb|ABA57136.1| Peptidase S24, S26A and S26B, 2.00E-12: 32% over 184 aa***	peptidase; phiE125 gp52-like	-	MKTIAQIRRENLAIAIKRARSAKFLAELADTSEAYLSQLKNAVP NKATGVPRFMGDDLARRIEQAIGEETGWMDFDRTDTSTASDD VFATSGPADISAEAQIAGSVIAWEEPTDLPSDPDRVWIDRYDYFI SAGKGLIQWEVREKSALPFSMSFFRGVGSNPKDCKLLIVRGDS MEPFLFNRDTIMIDTAK*
Bglu421Egp02	+	808	1092	284	Vibrio cholerae], gi|21885288|gb|AAL59694.1| unknown, 0.08: 40% over 55 aa***	hypothetical novel protein	-	MDLKTYLSLERGRLVKLAADIDAHASDVSAWANRKRAVPIPFG LPIEKATGGAVLRVEMFPEKVIRDVWPELIGADPPAPFGRAGGN ECGGGAR*
Bglu421Egp03	-	1402	1650	248	***no hits found***	hypothetical novel protein	-	MNDQYAVDNSEAVQLVITSLVMALHKQGTLQIRDFVTELGNQI DELDRSEPGRRVNPHERRYYDWFLALEQQAARDLGQNPG*
Bglu421Egp04	+	1927	2310	383	Xylella fastidiosa Ann-1], gi|71901483|ref|ZP_00683570.1| hypothetical protein XfasoDRAFT_1058, 7.00E-12: 42% over 92 aa***	hypothetical conserved phage protein	-	MQLDIPFYEGPEDALKAAVQALGGAKKIGPMLWPDKSTDTAA RLLLDCLNPARSEKLDLSQTMMVFRLARDAGYFAPFQWFALEI GFEAKPATKAEEENRITTVIERATATLSNALRMLDRMQNSQA*
Bglu421Egp05	+	2319	2549	230	***no hits found***	hypothetical novel protein	-	MKRRPASIPESMTERELAARSERLARTRSAYPIDHPQLSEISELL HRICEAGSLPVARWYAGDALALLRAFSVKEK*
Bglu421Egp06	+	2549	3547	998	Burkholderia pseudomallei K96243], gi|52210604|emb|CAH36587.1| hypothetical protein, 2.00E-29: 66% over 112 aa***Burkholderia vietnamiensis G4], gi|67546436|ref|ZP_00424350.1| hypothetical protein Bcep1808DRAFT_2641, 3.00E-13: 41% over 109 aa***	hypothetical conserved protein	-	MDWFRWWHGTVTDPKFQWVARKSGQKVASVLAVWAVVLET GSTATQGNADATRGNVASLQCEDWDVALGLEDGAVQSIYDAM EAKGLIEDGRIVAWDARQPKREDGGDPNTGALSSTERSRLHRE RKRREATHRNEVQRDATHATDLQRLDKSREEESKPTEANASVP SAADATPATPDAAEPVEPTGAAEPDGASGDLLGDGDSDADAG GQGARRKPAELPCPVERIVEAYHELMPLNPRVRVLNAKRRRAI AARWREAATLTCKPFGYSTVEAGVDAWCRFFRVCADSDFLTG KVKPKPGQPPFLADIDFLVSPDGFTKCLENKYHREAS*
Bglu421Egp07	+	3544	4782	1238	Mycobacteriophage CJW1], gi|29565978|ref|NP_817549.1| gp100, 1.00E-40: 48% over 173 aa***Ralstonia solanacearum], gi|17429000|emb|CAD15684.1| PROBABLE DNA-METHYLTRANSFERASE (DNA-MODIFICATION METHYLASE) PROTEIN, 9.00E-11: 48% over 62 aa***	DNA methyl-transferase	-	VSAITAGSATLHLGDCRDVLKTIADGSIDAVVTDPPYELGFMGR DWDRSGVANDVAVWAECLRVLKPGGHLLAFSGARTYHRMAC AIEDAGFELRDQIMWIYGSGFPKSKNLDGDWAGWGTALKPAH EPICVARKPLVGTVAANVIEHGTGALNIDACRVPTHTARPWREA RRNEASDEARNAYGQGLAGSMAVEDTTAGRWPANVIHDGSDE VLAAFPDAPGQLADTSTKPDRRKTQNTYGALRAGRAGEASAD RRYTNAGGTNFAATPGARRLDTGSAARFFYCAKASRADRHEG LADPGPQFKHGSTMRDAENLGAERNGNHHPTVKPTELMAYLC RLVTPPGGIVLDPFMGSGSTGKACIREGFGFVGIDLTSEYVEIGR ARIEHELGRRAAEAAERDRQPDLFQEPA*

Table 5. Continued

Bglu421 protein	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
Bglu421Egp08	+	4779	5753	974	[Bacteriophage phiE125], gi|17975219|ref|NP_536414.1| putative PAPS reductase/sulfotransferase gp57, 2.00E-144: 76% over 332 aa***[Burkholderia thailandensis E264], gi|83718172|ref|YP_439561.1| gp51, 7.00E-163: 85% over 321 aa***[Burkholderia dolosa AUO158], gi|84361722|ref|ZP_00986376.1| COG0175: 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase, 6.00E-157: 80% over 325 aa***	PAPS reductase; phiE125 gp57-like	-	MSDRQTLHVVSLSGGKDSTATLLVALELHGRDSVRAVFADTGN EHEATYAYALDYLPGALGITVDVVRADFADEFATKRANLARLA AGEPESSVYGNREFMYAWTPQAAARALDLLHPTGNPFLDLCM VRGGFPSRKRQFCTEYLKRNPLTEYAIGLIDAGFFVESWQGVR ADESDARRWLPHYEWRGGHYAVFRPILRWNVADVFEAHAAA GIEPNPLYREGMSRVGCMPCINAQKAEIREIARRFPEHVERVAA WERLVSEVCRPFSPVSFFHMGTTGHMGQATTVLQVVEWSKTT RGGRQYDLLADAEPATACSSAYGLCE*
Bglu421Egp09	+	5776	6567	791	[Bacteriophage phi1026b], gi|38707958|ref|NP_945099.1| gp68, 3.00E-65: 54% over 262 aa***Bacteriophage phiE125], gi|17975220|ref|NP_536415.1| putative chromosome partitioning protein, 1.00E-64: 54% over 262 aa***Bordetella bronchiseptica RB50], gi|33568259|emb|CAE32171.1| phage-related conserved hypothetical protein, 7.00E-70: 57% over 263 aa***	putative chromosome partitioning protein; phi1026bgp68/ phiE125gp58-like	-	MAKNSIDAYGAKGKGNLLDFDPDDLTLVVDEAHPLYDSRVHL PLDEAMVRNIDFQGVLQPIEVTKNPETGDIEVVTGRQRVKNCR EANRRRAERGEPKRLIKGFVRKLTLEERDRVLSAATASENAIRQ QETPISRAEKMARQLAYRSEEDVAVIFGCNVQTVRASLSLLECC ADVQKAVESGQISVTHAKVLAKLQPSEQRAKVKELIAAGDGV KGHAKARAQRAVMDGDAPRMRSRKTIEAELAASSGERAAAL RWVLGQE*
Bglu421Egp10	+	6573	6737	164	***no significant hits***	hypothetical novel protein	-	MNAVDQYFDDGLRAVPQSIESEQSVLGALMIDNDAVDRIGRVA RRALSIATSTG*
Bglu421Egp11	+	6753	7970	1217	[Burkholderia pseudomallei 406e], gi|83680829|ref|ZP_00942533.1| COG0305: Replicative DNA helicase, 2.00E-120: 56% over 386 aa***Burkholderia mallei ATCC 23344], gi|52427064|gb|AAU47657.1| replicative DNA helicase, 4.00E-100: 52% over 393 aa***Burkholderia thailandensis E264], gi|83718789|ref|YP_442704.1| replicative DNA helicase, 4.00E-100: 52% over 393 aa***Burkholderia dolosa AUO158], gi|84362019|ref|ZP_00986656.1| COG0305: Replicative DNA helicase, 6.00E-100: 52% over 393 aa***	replicative DNA helicase	-	MRLIVAGRSADVITVYESLQSHGDLDKTGGLAYLNSIVQNTPG SANIGRYAEIVVERAKLRMIMSAADEVSAEVQNRNGRTADELI AYAQSKFEPLSEGRTDGPKFIGTFLSPVVNQMDQEHQGNAPLA TTTGYRELDFKLGGGMRGGDLIVIAARPSMGKTALAQGIAEYV AQAQGTALIFSLEMPGTQLTQRAIARQGSIQLRNVRNGSHMTD ADWPKLTNVVGYLSDLPLMIDETPGLSMAEIASRARSVKRKH GLKLIVVDYLQLMTGGTDERHDLRIASYSGGLKMLAKQLDVP VIALSQLNRQLEQRPNKRPIMADLRDSGAIEQDADTIIFLYRDE VYHEDSPDAGTAEAIIAKQRNGALGTAYLAFVHEQAKFGDLA MGYYPQPRSAPVKSRGFNDD*

a total of eight proteins - one protein with putative function based on similarities to the database, four proteins with close resemblance to hypothetical proteins, and three that had no significant hits in the database. Contig B is the largest of the contigs, holding 21,545 bps in 29 predicted ORFs (open reading frames). 14 of Contig B's ORFs are similar to other hypothetical proteins in the database, eight had no hits or no significant hits in the database. Contig C's 8,027 bps encode for 14 proteins, six had hits to hypothetical conserved proteins, and five had no homologs in the database. Contig D is 7,483 bps long with eight predicted gene products. Five of the predicted proteins in Contig D are similar to hypothetical conserved proteins and only one had no homologs in the database.

Integration cassette. Predicted protein Bglu421gpA01 had a high degree of similarity (2.00E-81) to the site-specific integrase in phi1026b gp33 gi|38707923| as well as other integrases, strongly supporting a hypothetical function of integrase protein involved in the first stages of lysogenization.

Capsid assembly cassette. Through a highly significant similarity of 1.00E-118, the functionality of Contig B's capsid portal protein [05] was determined with respect to *Burkholderia pseudomallei* Pasteur gi|67756042| annotated as a bacteriophage capsid portal protein.

Tail assembly cassette. Contig B encodes for: a phage-tail sheath protein [17] determined by similarity to *Acidiphilium cryptum* JF-5 gi|88939387| phage tail sheath protein with a value of 2.00E-25; tail protein [24] identified by close resemblance to *Burkholderia pseudomallei* Pasteur gi|67756022| COG5283 phage-related tail protein, with an expect value of 5.00E-37; and a baseplate assembly protein V [27], functionality determined by resemblance to *Acidiphilium cryptum* JF-5 gi|88939381| phage baseplate assembly protein V (5.00E-19). *Lysis cassette.* Contig D appears to contain at least part of the lysis cassette with a lysozyme [06] whose functionality was determined by association with *Azoarcus* sp. EbN1 gi|56315596| putative lysozyme that are similar by a degree of 1.00E-35.

Morons. Bglu421 may seem to have sparse similarities between its proteins and other functional proteins, but resemblance to what seems to be morons are abundant.

Contig B contains: a predicted anti-repressor [19], determined by similarity to Escherichia coli E110019 gi|75235791| COG3617 prophage antirepressor, 6.00E-20; methionyl-tRNA synthetase [21], predicted as such by its resemblance to *Burkholderia pseudomallei* Pasteur gi|67756024| COG0143 methionyl-tRNA synthetase with an expect value of 2.00E-20.

Contig C also encodes for some interesting genes: metal-dependent amidase/aminoacylase/carboxypeptidase [01], functionality based on similarity degree of 5.00E-17 to *Burkholderia pseudomallei* Pasteur gi|67756051| COG1473 metal-dependent amidase/aminoacylase/carboxypeptidase – whose nomenclature indicates involvement in peptide modification; an ERF (essential recombination function) protein [07], predicted as such by close similarity to ERF protein in *Silicibacter* sp. TM1040 gi|69298208| with a resemblance of 5.00E-14; and delta 1-pyrroline-5-carboxylate dehydrogenase [11], functionality determined by similarity of 5.00E-33 to *Burkholderia dolosa* AUO158 gi|84361734| COG4230 delta 1-pyrroline-5-carboxylate dehydrogenase – chemical name suggests highly specific dehydrogenase activity at carboxy-termini.

Contig D also has an interesting protein: an exo-beta-1, 3-glucanase [04] determined by its similarity of 4.00E-05 to *Aspergillus fumigatus* Af293 gi|70988641| exo-beta-1, 3-glucanase - whose chemical name indicates a modification to structures in plant cell walls. Contig E also plays a part in harboring morons: peptidase-like protein [01] similar to *Nitrosococcus oceani* ATCC 19707 gi|76882455| peptidase S24, S26A and S26B by a degree of 2.00E-12 – most likely involved in breaking up peptides; a DNA methyl-transferase [07], functionality determined by resemblance (9.00E-11) to *Ralstonia solanacearum* gi|17429000|emb|CAD15684.1| probable DNA-methyltransferase (DNA-modification methylase) protein which, as annotated, seems to be involved in DNA-modification so that the host won't recognize its DNA; a 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS) [08] determined by its striking similarity of 2.00E-144 to Bacteriophage phiE125 gi|17975219| putative PAPS

reductase/sulfotransferase gp57 – which seems to reduce the PAPS molecule found in plants; Bglu421Egp09, functionality determined as a putative chromosome partitioning protein SpoJ by its similarity of 1.00E-64 to phiE125gp58 gi|17975220| putative chromosome partitioning protein SpoJ; and a replicative DNA helicase [11], annotated as such by its high degree of similarity (2.00E-120) to *Burkholderia pseudomallei* 406e gi|83680829| COG0305 replicative DNA helicase.

Identification of a Bglu421-related prophage in *Burkholderia pseudomallei* Pasteur. In performing Blasts of Bglu421 contigs, it was observed that 21 of its 70 proteins were clustered to homologs in *B. pseudomallei* Pasteur. TIGR (The Institute for Genomic Research) has recently deposited the genomic sequence of *B. pseudomallei* Pasteur in its entirety, however this genome has not yet been manually edited for genomic organization and structure. Its clustered protein resemblance to Bglu421 indicates that an unannotated prophage exists between proteins gi|67756015 (nt 2,217,234) and gi|67756053 (nt 2,217,638) of *B. pseudomallei* Pasteur. Because of the rough quality annotation of the Pasteur sequence, the relevant region most likely containing the phage had to be first extracted from the database and then re-edited on Artemis v.7. The re-annotation process involved assessing the strength of Shine-Dalgarno sequences and stop codons leading to necessary modification of ORFs and addition of novel genes. This putative prophage will be referred to as phiPasteur.

Genome overview. After editing and some re-annotation, phiPasteur was found to contain 45 proteins over 32,639 bps. A summary of predicted proteins and new additions are listed in Table 6. Note that newly added ORFs are designated by letters and unchanged, pre-existing ones are designated by numbers.

Capsid assembly cassette. phiPasteur contains a capsid portal protein [26] whose function was identified by TIGR's annotation.

Tail assembly cassette. Re-annotation of phiPasteur revealed three new tail assembly cassette components: a tail fiber [D], highly similar (2.00E-56) to that of *Burkholderia cenocepacia* phage BcepMu gi|48696962|ref|YP_024725.1| gp52; a tail protein [G], whose function was determined by close resemblance (5.00E-04)

TABLE 6. Summary of phiPasteur coding regions and edited proteins

gene id Genbank	Strand	Right end	Left end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
gi|67756015	-	561	1	560	SELF	hypothetical conserved (as EF-1a in Genbank)	-	MTTMKTHETRRTLAEDVFYPDHPPRTESVVFRKTKHEGHAVQLPC AISGHTEGVEYHHLMCEWAFSGAVDWRTVKGIATGEISRLPVIDLE TDQPTGDTFDARQSLLWALCKVAEVRGFDWSSFDPSRPETFVDSM ANMLVLHSKFHRGHGHGIHELTLPVWIFQALPRNPGFIFSPDELHN QLTH*
new					***no hits found***	hypothetical novel protein	-	MKRGARLPRTFSTRTIRRVPSRWCSEKRNTKGTPCNCRARSAGTPR VSNITTSCANGRSAVRSIGAP*
extended start	-	962	558	404	[*Burkholderia pseudomallei* Pasteur], gi|67756016|ref|ZP_00494904.1| hypothetical protein BpseP_01001684, 134, 1E-73: over 134aa*** [*Burkholderia cepacia* phage Bcep176], gi|77864683|ref|YP_355393.1|gp58, 165, 9.00E-54:66% over 116aa*** [*Burkholderia dolosa* AUO158], gi|84361689|ref|ZP_00986343.1| , COG3772: Phage-related lysozyme (muraminidase), 153, 4.00E-50:62% over 100aa***	SAR lysozyme; Bcep176 gp58/ phi1026b gp24/phiE125 gp25	+	MANVPKKTLAGVVGAAAAALLFSVVPKFEGVKLVGYLDPVGIPT KCMGDTRDVVVGRAYSEAECRSSLETQLIAHAEPVLRCTPGLRGR PYQLAAAVSFAYNVGAHAYCNSTTAKRFNAGDLRGACRAINESDS GRPQWVFANCRTVIDPKTKKPLPVCDTLPGLVKRRAEERALCERD L*
gi|67756017	-	1275	1093	182	SELF	hypothetical conserved protein	+	MKNDLAVSAAKAAPAVGSNFWLWLTGHDINWWVAVATIAYIGLC AYYLIKNKGKRALLDG*
different reading frame, extended end, including gi|67756019	+	1733	1449	284	SELF - different reading frame, extended end, encompassing gi|67756018	hypothetical conserved protein	-	MKLAFRVRSTLACPHAHQVHTSRNAGSSDAQARAEQDRQADGR SILAGRRSSVAQVHRGAVAAGAEPSHGVRCSSEARAHGGRVRSRP REDGDLRRARRGGARWRAAAIAMALGCTAALTACGGGSDGGGQ TAKPTTHQDPMPAPSTPHSKIVKIAMYGDSTAFGTTLLSGVYVQSQ HNEPASLQFALQNKYGLAVTVENRGVPGSTCGQWLWGQKDVRQ AWTVEMATSDAQIVMMNCAINDAFLPNETDQDFQYVYGQFAQIA RQYGKTFVILTPNPIDDPHNDRLEELVRNQHYVAQVQQVKIIDQW SVIQQAMPNWRASLPDRIHPNDALYEYMAQITAAALDSTLGG*
new	+	3595	2525	1070	[*Mesorhizobium loti* MAFF303099], gi|14026868|dbj|BAB53464.1| acetyltransferase , 357, 2.00E-51:39% over 333aa***[*Agrobacterium tumefaciens* str. C58], gi|17743099|gb|AAL45401.1| sugar acetylase , 372 4.00E-48:38% over *319aa****[*Xanthomonas oryzae* pv. oryzae MAFF , gi|84366310|dbj|BAE67468.1| putative o-antigen acetylase , 350 2.00E-39:34% over 344aa***	o-antigen acetylase	+	MHRNNFDFLRLLAALLVVVGHAYGIMAHEQPTLLGPQISVLGLIIF FSISGYLVTNSWKSDPSLGRFLLKRALRIFPGLIAAISVTTFVIGPIAS DLPLASYLTQRGTYAYLLNGALRYTYGLPGVFADNPISGAMNISL WSLPVEFSLYLAVPVIVLLGARRYDVAFWAVTVVLAISSIYLTYYQ HGPHIVIYGTDLTAACGIAYFFTAGAYFAAHKGKFSMVVACALFAA WAIEPRFFTGILRFPIAVTVAFAIPYIILSIGTKSWPIIRHTGRLGDLSY GIYLYAFPVQQIIAQGALRKASVSTSIWLSLLITIPLALISWHLIEKQ AIRAKGMILNRLSGMASPAERAPVVR*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgpD	new	-	3732	5819	2087	*Burkholderia cenocepacia* phage BcepMu], gi|48696962|ref|YP_024725.1| gp52 , 785, 2.00E-56:36% over 520aa***[*Xylella fastidiosa* 9a5c], gi|9107487|gb|AAF85117.1| hypothetical protein XF2318, 277, 1e-78: 61% over 148aa***	BcepMu gp52 tail fiber	-	MHDRVPLKAPQSLGFLFFWSPFLMDRQTVYSGQVPLETDLLNTNRN VLVALGKVLGDMLGTSTFATGLGCVPTAPASMSVQVQPGQLYSLQNI DNTAYSSLAADTTHQIVKQGIALDAQSFACAAPTTAGYSINYLIEANY SDVDTNAIVLPFYNASNPSQAFNGPGGNGQPSMTTRAGKVILTLKPG VAAATGAQITPAVDAGCVGLWVVVTVAYGQSTITAGNISQYTGAPILPT SLLQSLITNNLEYGIDAGPANTVQATFPIAPMAVVDNQQFWVKIKSSN TGATTFTPNPGVIGALPVVGAAHQALQGGELVANGRALFVYRSDIAS YVLESCTGAAMQTAPATQSGQAVQKSQLGNFQAFYAYTSSQTLTASQ AGSVINFWGGSPSTFSLPACSSMAAGGAFLINNSGAAVLTVTRSGTDS ILCNGGSASIVVSPGDNLLVFVIPPSQWIATGGSAQLPFAGTMQRVSG GVVGAARNAYMNVSAASASATYTADEVVVETVLGGTFFRLAGFNK AINLTTTGAGGMDTGAAPTSGYVALYAIYNATTQASALLATNATSAV PSVYGGAYMPAGYTASALVSVWPTNGSGQFVAAVQADREISIPYVTV LTTSTPQASPTALSISAAVPPNARKASGTMSVSSTSSTPNSSLSLYATAS GAGIQGLNNSVSASGGISTNYRGVPLSTSQSLYYSATSSAGTPTFSISV ASYEI*
phiPasteurgpE	new	-	5893	6558	665	[*Xylella fastidiosa* 9a5c], gi|9107484|gb|AAF85116.1| hypothetical protein XF_2317 , 220, 6.00E-73:61% over 221aa***[*Xylella fastidiosa* Ann-1], gi|71898334|ref|ZP_00680507.1| conserved hypothetical protein , 194, 4.00E-65:63% over 192aa***plus 8 others	hypothetical conserved protein	-	MATGDQNDILARLTRLIPVSWFESFHAIRDAIVSGLANAHSYVYSFIT YVGLQTRILTATDGFLDAIAQDFFGASLLRAANQTDASFRARIIINIFRE RATRNAIIKVLKDLTGRTPVVIEPQRPADTGAYGAPNIGYGVAGAYGS VLLDYQAFVIAYRPIGTGIPYIAGYGSSPGGYRQPSRASYADLSLIQNT VSDADIYAAIDSVKPAGTIIWTSISS*
phiPasteurgpF	new	-	6566	7693	1127	[*Acidiphilium cryptum* JF-5], gi|88939380|ref|ZP_01144829.1| putative phage protein , 374, 5.00E-72:48% over 346aa***[*Gluconobacter oxydans* 621H], gi|58002093|gb|AAW60987.1| Putative phage protein , 406, 2.00E-45:37% over 351aa***plus 5 others	hypothetical conserved phage protein	-	MTFQTQTFTQVVENAVVAVQSAASNLIDTAVGSVLRAILEANAAIAM WLQGLILYLASITRAATCSGSDLDSWMADYGFTRLPAAYASGQVTFS RFTATNQAVIAVGTLVQSADGTQQYSVIADTTNPAYSAALNAIVIPAG TSSVTATVQAAVAGAAGNMLANQVNTLSQAIPGVDTVTNANPFTNG ADAEADAVFRTRFQAYIASLSKGTKGAVAYAITSIQQGVSYSFTENQT YGGVTNMGFFYVVVDDGSGSPPSSFLSSVYNAIDAVRPVTSTFAVFAP ALETANVGMALTTTGNSATHNAVIAIVTTALQAYINALPIGVSLPFSRL AQIAYDASPLVTNVTAVTLNGGTSDLTATSQQIIKSGTLVIV*
phiPasteurgpG	new	-	7757	8161	404	[*Gluconobacter oxydans* 621H], gi|58002092|gb|AAW60986.1| Putative phage tail protein , 121, 5.00E-04:35% over 78aa***[*Gluconobacter oxydans* 621H], gi|58003269|gb|AAW62163.1| Hypothetical protein GOX2432 , 119, 5.00E-05:38% over 68aa***	hypothetical phage tail protein	-	MLNDLNHYFGNDLVASSGGDLGLASGGLRGQQRVLRRLLTNPALKD SAGNVIAPGDYIFHPDYGAGVRRMVGSNAGVDEIRAVILSQMLLESS VARTPPPAVSVLQIVGGVSVTIQYVDADGGGNVVLAFDVTN*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgpH	new	-	8158	8823	665	[*Acidiphilium cryptum* JF-5], gi|88939381|ref|ZP_01144830.1| Phage baseplate assembly protein V , 182, 2.00E-25:34% over 202aa***[*Gluconobacter oxydans* , gi|58002091|gb|AAW60985.1| Phage-related baseplate assembly protein , 207, 7.00E-08:31% over 97aa***	baseplate assembly protein V; Acidiphilium cryptum JF-5-like	-	MSLHLKALASAVRGQAAMERGNTAHSRYALVKSVDPNSYSVKVELQ PDRIMTGWIPVKSQWVGSGWGLASLPTVDQLVVVDPIDGFNDAWIV SGYLYNDSERPPSSAVGEIRMVHQSGSYIRLTNDGKVGVNGQAEVDV SGPTINIQATGNVAVNAGGAASVTAPSIQLGAAGQTLLAFVTSAFMSL FNGHTHNDPQGGVSEAPNQQMSSAHLTSTVKGG*
phiPasteurgp04	gi|67756020	-	8820	9950	1130	SELF	hypothetical protein	-	MLNTISPIGVSRQPRGVVKLNGVPIDAWIDFEVDNNGFYQADTFRAN FAANWLPLDRNKAWLLSQADLTLELFAGFPADPQNYDSSALKSLIIGR VDEVQFRPADGVIELSGRDYTSQFLDAKTAEQFSNKTSSQVAQILAQR HNMTAVVQATTTPIGTYYQIDHVRVSDSRPEWDLLTWLAHKENFAC YVEGTTLYFQPLADPSKANTYVIQWKDPTGPSASPMANVIDIDFSRNL TIAKDVSVIVRSFNAKQKKGFTVTAKAQHNKNKVIRNAQLPYGQPQ VYSYFIPNLTQAEAQARANALLQDISKHEMRFRANMPADNILTPRTLV SVTGTGTAFDQIYYPDSIIRTMSIRDGYRMTLSAKNHSVESVPTP*
phiPasteurgp05	gi|67756021	-	9955	10671	716	SELF	hypothetical protein	-	MLTWGEFSYIVLIEHFKPVYERFYLVPYTITLEVVQNNTNAITTIALPT IDDAVASDMAGASSLVSALGIPSLTSAFGTLQSAISSVNSFVGATRSVV NGVLQPINQIRGQITSLRTQFDALTAAQGLLGKISVGIGLGTQASSLLS TISSVQSSTTLFNLDKLIGRVQANVSSVFSANKTETIGGGDLFTLASKH YDDPSAWTYIANANGLTDPSLSGINTLTIPPAPNISGGILQS*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgp06	gi|67756022	-	10911	12767	1856	SELF	tail protein	+	MLAGGALTATGIFGLKVLETPLDKATQYEKYVAKMRQMGLGDAQIQ DAKKFVQANEVIGTSMLDRMRLFSEAQGSFRESGMSGLKALDAAKT MTPVLATYEVAMATLDDKSHAKAEGAMRNLNKTIEMMGGLGDTKR AQEIADGVFKASQSSGGMVDERHLKQFFAYGSSASNQQLLRTVFGGL EPIIGELGGSTVGTGMRTAYNRVNGTMALVPKNVRRELKRLGMARD ATGMEQTDALARLQASDVIGYVQDIMRRYEKAGIKTQIDRERENSIL FGTNGAKIYNKLMSQMSVIEESLHAYDKAQGASQVVNDPANQKLQA QRRFEKKIEDLQLTIGRDGGLLDMLNRGLTFLGGAIERVTKFAEGHPK LTKFAVGGFAFLSMVTIIAGEVLLLGGAFRALSTAIAFTRVGSTMVRIA QGSRLALMALRPVGNGLVFLGGIVRGLLGPFALLARGLLILGRALAM GAIGGIGGIIRLAGALAGTATTALSGALAALVSPLGIVVLALGTLAAA AYAFSPISQKEIDSYKTDGGVKLTASAAARLAAVDSGATTNGPNIGPR QQSVVHLQHQTLLDGRVIADSAAKYIVGSMNAPVGTGAVDTRQTMP AVGSPYPFR*
phiPasteurgp07	gi|67756023	-	13172	13687	515	SELF	hypothetical protein	-	MREFNSLTAFAAHLIRLDGGILLRMHAGLDRVAAHIERTAKAEIGQYQ PAVGPFDAWAPLAEATKDDRVRNGFPEDEPLLRTGEMRDSISRRVRG EAEIGSTSDKLVYQELGTANAAHPIPPRPVLGPAAVLAEPVIRRVLIGA LVSGLLDGGSMPAGLIADYESATVTR*
phiPasteurgp08	gi|67756024	-	13854	14267	413	SELF	methionyl-tRNA synthetase	-	MTTTTIKDTTISPGAAQSNVETPTQAVMRQTGQEVTIEDSRGRKIALR KPGVLAQYAIIDAVGGESAKNPVYMSMVLPLIYVASIDGEAVSTPNTK LEVKALIKRLDEDGIEAVISAVSEKFAAPDPDADRETLKKS*
phiPasteurgp09	gi|67756025	-	14345	14776	431	SELF	hypothetical protein	-	MSINNFSVGRDISLDIVTGSGPLQINLITEFTSNQESTEVKVKGIDGITR FVRFFDGWKGKFNVERQDATLDQYFSQLEQNFFSGIGEQSVTITESIQ EANGSISQFRYLGVLLKYDDAGGWKGDSTVKQAVSFVCSRRIQIA*
phiPasteurgp10	gi|67756026	-	14835	16349	1514	SELF	hypothetical protein	-	MVQAGSINTTALIVPDVYVQIVPPSVALLNGLPTNILGAVGTATWGPV NSPTTVGSMADYAQKFGAIQARKYDMGTFVAAAVLQGANNFRCVR VTDGTDVAASTTITAGTAAVATAIAAAINNGLSVMRGASNLAVASSST TIVTVTAKYTGSLGNNVQVTTSAGSAANTTKVTIALPGLTPEVFDNIP GAATAGTTTMVGGTDGATTITGTTLVGVDTVPRKGMYCLRNTFTSIA ALVDCDDSTTWATQVAFGLSEGVYMIGTGPAGDTIANATAAKATAGI DSYAFKLLFGDWVYFLDTVNGVTRVISPQGFIAGLLANLSPQNSSLN KPLYGVVATQKSILNQTYSAAELQSLGQAGIDVVTNPVPGGSYFGAR FGHNSSSNAVINGDNYTRMTNYIAYSLNAGMGKFVGQLQSLTVQRQ ARSTISTSLDNMWGQGMIGNSLGTVPYSVQLDAGNNPQSRVALGYM QADVKVQYLSVIEKFLINVEGGQSVQIVRQSTALA*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgp11	gi|67756027	-	16424	16606	182	SELF	hypothetical protein	-	MLGLVVTHAFGTWEKGDLIADPAFVDEVLASEHAERVVKVNVPDPV DTTAPRSRSAKKAP*
phiPasteurgp12	gi|67756028	-	16618	17466	848	SELF	hypothetical protein	-	MADLSDVGNVLAGLIGGFIYPNGTGAPSAAGLPVKVYPGWPDPASL DADMAAGIAHVSIFPHPTGRNTTRYSGDWKTTAAPAQTLAASVNGN AITISGTVATPQTVAVIANGAPFVYAAQASDTLNSIAAALATLIAAQIP(TTSSGAVIALPNSARIAAARISASGTIARVVRSQEHSVQVTVWAPSPDT RIHVATAFDALLSDLRFLSMPDGFGARFVFSNSYDTDRLAKSTAYRRD VLYSVDFATTIMQQAAQIAVEQLNVTPQVSGVDGSSPVFTTYE*
phiPasteurgp13	gi|67756029	-	17459	17659	200	SELF	hypothetical protein	-	MEPKQITIESDGTALGTKVFDAAGKPMDMSSITRIDWSLDGCGAVAE ARITFAFPKVNVRGALNDG*
phiPasteurgp14	gi|67756030	-	17659	18003	344	SELF	hypothetical protein	-	MTDAIYNRTISIFRQAAQTGVGTIGYGGTTPAQQAQIATGISASIQLVK SKAANPAGLPTDVSTTEWRIFYRSTNPSLMSERDVIVDDLGRRFEVM GAYPNAFGYSALCKLLEL*
phiPasteurgp15	gi|67756031	-	1800	18911	17111	SELF	hypothetical protein	-	MPSQYLQSADFATYGVPNATAAQVTQASTLIDAFLQRPDGLLWAPDA NGVPAYMAAKNADVSFALSAAIAPGNDVPVSLPMPAAQLDVQAGD VVLLDRASTTACEAVVIDGIVGNTVRLRNVQFAHAQGATLEFGMVIT QQCYVPTQRPVVRLQNTPVFRVIGGAGRFGYGRRGDAGASLMDDY NLLASLTQFGGPPLWEVFDPTVGSLDRGTGQFWLPSSVYLAYYSEVR VHYVAGYSYASLPGAVKSACASLITAASNSSLFAGGVKTYRAGDTEIE RFANTVFDDDTKRSLSPLRPMEFV*
phiPasteurgp16	gi|67756032	-	18915	19211	296	SELF	hypothetical protein	-	MSVALQILIGAQTGSANSNPFQQSVADGLCTIVAGGLASSETATVQIQ DASGNWQNVPAAIAPQLTPTVPAAVLTAPGVYRVVKSATVGSVGVAI YRV*
phiPasteurgp17	gi|67756033	-	19208	19435	227	SELF	hypothetical protein	-	MKIYLPGGKGKHTIYSQPGREFPTSDFCEADGRPKMFATVFVEGIAD VTAELGRYLVDQGVASRSPLIVPQGIAA*
phiPasteurgp18	gi|67756034	-	19559	20527	968	SELF	hypothetical protein	-	MDLSAAADYLGTGAIEIPIFEREIMDVVRRTSVALNRIEQVPATGHPHR YFEQTAIGQAAAVDPRNLSSTATGPTRVERPAFIKAVTNQSNLSLFDK DVTEQQGQFASVVARDVDDIINAIEIKRASMLWVGTDTSLSAPTTLE WMGGLAQITQTATVAPGASIIDALKAQVAAMVANVTYVVRPTAVYL NPVLADYIDQEAKASRITLDSMEVTAGVTVTAISTQAGKLPLIGDPFM PADTAAKYGFSAPPAGNKNYYAVIMSEKDVEIPVISGREFNKNPRLFQ LGLTGNLAGQFVGVKFDAVIFKGASYAHTVLAVQRP*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgp19	gi|67756035	-	20645	21991	1346	SELF	methyl-accepting chemotaxis protein	-	MVDRTQGLTPDERRTARERIKRRAHQIGMSTKGWKADVSASVEICA QYELAAMSIAVPDPDHPNKMPFTGVLTRVDEPSDNPPGGSDGHRTFIP AEVAEAALGSLLGMCVDFTPNFDGHDKTAKIGLITEAYIDGSALRIGG FFYANDFPNETAEIKAKKDRLGFSYEAKAAIRDVNAETWHVDRIVFT GAAVLLKDKAAYTTTSLEAARANQEPTMEELKQLMEAVGALTTKVT EIAASQAATSEAVSKLNEKAIEAAAARDKVRPHAESLRSCAASMEAA GIGTHSTSGHANVLRHMAASMEAEAAMGNIPHIYRDHDYLSRVMEA GAAGAAGAGAAVGADKGAKADDATAKQIAELAASVASMTTVVTDL KAAAFNGAREPERKTLSPEMKTLLAKAGLTDQAEKGTLTVSEVDKT LEAAGITGHAAMTAKLKLMHSGILPVGKQ*
phiPasteurgp20	gi|67756036	-	22115	22588	473	SELF	hypothetical protein	-	MAKQSNAPTAQTKQAQGGAGDGKDVDLNTGAASKLPEGDRKMPA DAGAPPKLPVDDAAKQDGAATASRAPVAGLDFAAEGGDKTVASMFI ATTLAGHVSNLLTSAETAGDTASQSVMNRIHVKLAEFRHMLTSIEQH VLDAAHADVQELVDDMKRLF*
phiPasteurgp21	gi|67756037	-	22620	22811	191	SELF	hypothetical protein	-	MATNQSTLTARRIALEQSKNPQVAGLIQAAEVAAADLRNRLNALAAT GDSAADTAIAAIKAAL*
phiPasteurgp22	gi|67756038	-	22821	23126	305	SELF	hypothetical protein	+	MSFDSHTVSAPAGAKPQLNDRAVLGSLIGVANVAGAGAGAAVTVAV TGLSLPPNYAVAVNPGQDATWFVSAKTQSGFTVTLNPRLAANTLTAG TIDVIITA*
phiPasteurgp23	gi|67756039	-	23189	23461	272	SELF	hypothetical protein	-	MTNPPDWVELNRDRCEAIAAAVYRGACTTAHTHQRALAASLRVFAE ARDSIFVEHHFEKYPTERMVRQIPRALYEELTGELTDEAFASAH*
phiPasteurgp24	gi|67756040	-	23645	23869	224	SELF	hypothetical protein	-	MDNQHKQIKGYRDLTAGEISLMNDVKHRGEKLHELIVNLRTSPGIDQ RWVDIAETQLQLGLMALTRAIAQPTSF*
phiPasteurgp25	gi|67756041	-	24476	24760	284	SELF	hypothetical protein	-	MTEAECLREVERDPWAHVEPSHYTARPHFESFRNGYGDPRWVVNGT MYEHPTVNGGTHCVWEWPDGQAIEFWRDAARRVVAGPRVTQRTKP GAG*
phiPasteurgp26	gi|67756042	-	24925	26223	1298	SELF	capsid portal protein	-	MLFAARRRLQPKRQTQAYARLMNIGGQQYQRDRPLLKPTPSNLRNF ARTPFARRAINTIKNPIAELRWDVRPKEGVTPSRVLDQQLRVVSDCLS RPNNDDSFTSFVEQEVSDLLICGAFAYEHSLSGDPTRPLWLWPTDSMS IQIFAGWDGNPKSTRYYQTLGAGNVGGAQGLPIAAQDLVYGRGDPT TDTPFSWGALEIAFQTINRLLGVSAYAGNVASNATPSTLLNFGSAGKA GQSPALLDPNKLRAFRDWWRNDIEGQGMTPITSFEAVQAVNLRGSD DKALFLEYQDMLKREIAAAFGISPMNLAIERDVNRDTAEVSYDRDW DGTIKPTAHLLRAHINRNSIEGRLGFSQIELSFPDIDREDEKANADIYK VEYESNATTPNQYRARKGLEPLESAWGDLTYADVQIAVKGAQGAKS MNPDLATKE*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgp27	gi|67756043	-	26271	27698	1427	SELF	hypothetical phage protein	-	MITDPVELQAIRDLAQEDLYFFSRYMFYRRRNFKWKRNWHHKAICD ALMAVYRGEIRNLIINIPPRYSKTELAVVNFCAWAMGKAPDSEFIHTS YSAKLAAKNSAGVKEQLTHEAFADVFPGVLLSADSAAKDDWKTTA GGSFYATGSGGTITGFGAGKARDGFAGAIVIDDPHKADEASSDTIREG VIEWFQNTLESRVNNPLTPKILIMQRLHERDLAGWLLDGGNGEEWAH LCIPSINDGSWPEWSGGKVGDALWPWKHSIDALRVMERKKPYVFSG QYQQKPSPSEGGIFKPDRITIVDALPFDLTQARGWDFAATTETEAADP DYTVGAKLATDSIGRIFIVDIVRDRMTPFDVEMSLKNTAARDGNNVR ISIPQDPGQAGKMQSLYFVQKLAGYIVSATTETGAKETRAAPLASQV EAGNVYMLRAPWNDALIAEMRMFPNGAHDDQVDACSRSFSALFSQ SSPAWTSA*
phiPasteurgp28	gi|67756044	-	27688	28173	485	SELF	hypothetical protein	-	MAKLTDDVKAFIVRALACYDTPEQVARAVKQEFGLDVSRQACEAYD PNKYVGRNLSQKWRVIFEKTREAFKTDVGDIPIANISFRLRTLQRTLA EAEKRGNTVVVLQALEQAAKEAGGAFTNKTRVEASGPNGAPLQHG VVTTTVPPDEYKKVVRQALDDY*
phiPasteurgp29	gi|67756045	+	28966	29121	155	SELF	hypothetical protein	-	MQHYIVVMENAAMGRTNVHMHAVSAEAARDNVVEYWLSMGEVA QCVAVYAD*
phiPasteurgp30	gi|67756046	+	29148	29459	311	SELF	hypothetical protein	-	MALDHGLLNLPLAKRGDIDAQIDAYKREQAAAEKSARKARAERLKL DKAAAAPVLAQLLADVELINRKAAEMNVTPKTLRDQLKSWATWQP KNLIALGAKWLA*
phiPasteurgp31	gi|67756047	+	29489	29644	155	SELF	hypothetical protein	-	MEKLYSEWFWSEASARGAAVRAAKKVGGVARWRYAMRADGQHD WIAEVFGA*
phiPasteurgp32	gi|67756048	+	29673	29852	179	SELF	heliscase subunit of the DNA excision repair complex	-	MDETAKLKADIAKLRKALREARTYVDDFGRNDNGELTRAAKRTFDS ISLALYETDGRRA*
phiPasteurgp33	gi|67756049	+	30563	30829	266	S*ELF, [Burkholderia cepacia* phage Bcep176], gi|77864664|ref|YP_355374.1| gp39, 105, 1e-14:63% over 43aa***	hypothetical protein; Bcep176 gp39-like	-	MNEITTGPELEACGWYVRTKRTDVDAAGLLVADCSIHPYGAEYARLF AAAIEMREALQIVERELGKGFTPMLHLPRIRAVLAKAAPAT*
phiPasteurgp34	gi|67756050	+	30920	31123	203	SELF	hypothetical protein	-	MDDAIKDQMTGFLDRLLLGTALDEVGNHDDDSLLEISSRCAMRGDN YTRRKLEGIHDAARQLRKRFQ*
phiPasteurgp35	gi|67756051	+	31162	31416	254	SELF	Metal-dependent amidase/aminoacyla se/carboxypeptidase	-	MSTTATAIPLREVESSQIHAIGHDAETQTLAIQFKSKAGAGSIYHYQNF NAAEFELFSSAESIGSYFGKYIKPFPEIYPLEKIG*

Table 6. Continued

phiPasteur protein	gene id Genbank	Strand	Left end	Right end	AA length	Representative BLAST hits (Query Genome, Gene, AA length, Expect: % identities overlap length)	Protein description	THM domain	AA sequence
phiPasteurgp36	gi|67756052	+	31613	31963	350	SELF	hypothetical protein	-	MRNNAIYGRDYGEWPWAFLCRGCGSYVGMHPFTDIPLGTLADAATR EARKRAKAAFNPLWQSGSMTRTEAYTWLAARLGVAVAACHIGWFDV AMCQLVEHECVTTPAWARTRPTSG*
phiPasteurgp37	gi|67756053	-	31920	32639	719	SELF	Diadenosine tetraphosphatase and related serine/threonine	-	MPQVERFPANVAGRDFIVGDVHGCVAQLVILLQDVKFDRAVDRLFSV GDLIDRGPDSPSALALLYEPWFHAVRGNHEQMLATVIDQPTARHWD WWTQNGGAWARCERAATLRLYADLVNALPLAIVVGDGPERFNVMH GEFFGTDDGLDSGDFAPHVAMAMLWGRDLINGKAVPPREGLSVTYV GHTPVRTPFCAFSHHYLDTGAGHRTRDARLTLYQHGVGEAASIRTW DAFAPTLAW*

Gluconobacter oxydans 621H gi|58002092| putative phage tail protein; baseplate assembly protein V [H] annotated as such due to its degree of similarity 2.00E-25 *Acidiphilium cryptum* JF-5 gi|88939381| Phage baseplate assembly protein V. Through TIGR's annotation, a tail protein [06] was also identified.

Morons. phiPasteur contains several morons pertinent to host fitness: o-antigen acetylase [C] determined by similarity (2.00E-39) to *Xanthomonas oryzae* pv. oryzae MAFF gi|84366310|dbj|BAE67468.1| putative o-antigen acetylase – which as 'antigen' suggests may play a role in modifying the o-antigen produced by the host of *B. pseudomallei* Pasteur in order to affect proper host identification of the bacterium through the antigen; methionyl-tRNA synthetase [08], methyl-accepting chemotaxis protein [19] metal-dependent amidase/aminoacylase/carboxypeptidase [35], a helicase subunit of DNA excision repair complex [32] diadenosine tetraphosphatase and related serine/threonine [37] – all originally annotated by TIGR.

Comparison of Bglu421 and phiPasteur. 22 of the 70 predicted proteins of Bglu421 had a counterpart in phiPasteur (see Table 7). 16 of these similar proteins were clustered in Bglu421's Contig B, with three others in adjacent Contig C and another three in neighboring Contig D. Once the observation was made that there was a high-level of clustering between proteins of Bglu421 and phiPasteur, the Bglu421 contigs were assembled against the predicted phiPasteur genome as a preliminary method of contig organization, resulting in the following sequence: Contig C, Contig B, Contig D, Contig A, Contig E. After the assembly, it was evident that there is indeed a consistent sequence of homologous proteins that runs anti-parallel to those in phiPasteur.

In Bglu421 and phiPasteur it is clear that there are clustered groups of consecutive hypothetical conserved proteins that are homologs of each other: Bglu421Bgp05 and phiPasteurgp26; Bglu421Bgp06 and phiPasteurgp22; Bglu421Bgp10 and phiPasteurgp18; Bglu421Bgp11 and phiPasteurgp17; Bglu421Cgp12 and phiPasteurgp36; Bglu421Bgp13 and phiPasteurgp15; Bglu421Bgp14 and phiPasteurgp14; Bglu421Bgp15 and phiPasteurgp12; Bglu421Bgp20 and phiPasteurgp09; Bglu421Bgp22 and phiPasteurgp07;

Bglu421Bgp25 and phiPasteurgp05; Bglu421Bgp26 and phiPasteurgp04; Bglu421Dgp07 and phiPasteurgp01. The pairs just listed had a range of degree of similarity from 8.00E-8 to 1.00E-118. Note that the gp numbers of phiPasteur are listed

TABLE 7. Bglu421 proteins to phiPasteur mapped proteins

Protein in Bglu421	Protein in phiPasteur	gi from Genbank	Encoded protein
Bglu421Bgp05	phiPasteurgp26	gi|67756042	hypothetical conserved protein
Bglu421Bgp06	phiPasteurgp22	gi|67756038	hypothetical conserved protein
Bglu421Bgp09	phiPasteurgp19	gi|67756035	methyl-accepting chemotaxis protein
Bglu421Bgp10	phiPasteurgp18	gi|67756034	hypothetical conserved protein
Bglu421Bgp11	phiPasteurgp17	gi|67756033	hypothetical conserved protein
Bglu421Bgp13	phiPasteurgp15	gi|67756031	hypothetical conserved protein
Bglu421Bgp14	phiPasteurgp14	gi|67756030	hypothetical conserved protein
Bglu421Bgp15	phiPasteurgp12	gi|67756028	hypothetical conserved protein
Bglu421Bgp17	phiPasteurgp10	gi|67756026	phage tail sheath protein
Bglu421Bgp20	phiPasteurgp09	gi|67756025	hypothetical conserved protein
Bglu421Bgp21	phiPasteurgp08	gi|67756024	methionyl-tRNA synthetase
Bglu421Bgp22	phiPasteurgp07	gi|67756023	hypothetical conserved protein
Bglu421Bgp24	phiPasteurgp06	gi|67756022	tail protein
Bglu421Bgp25	phiPasteurgp05	gi|67756021	hypothetical conserved protein
Bglu421Bgp26	phiPasteurgp04	gi|67756020	hypothetical conserved protein
Bglu421Bgp27	phiPasteurgpH	[*Acidiphilium cryptum* JF-5], gi|88939381|	baseplate assembly protein V
Bglu421Cgp01	phiPasteurgp35	gi|67756051	Metal-dependent amidase/aminoacylase/carboxypeptidase
Bglu421Cgp12	phiPasteurgp36	gi|67756052	hypothetical conserved protein
Bglu421Dgp01	phiPasteurgpE	[*Xylella fastidiosa* 9a5c], gi|9107484|	hypothetical conserved protein
Bglu421Dgp02	phiPasteurgpD	[*Xylella fastidiosa* 9a5c], gi|9107487	hypothetical conserved protein
Bglu421Dgp07	phiPasteurgp01	gi|67756015	hypothetical conserved protein

in relation to the map numbers and not to the current gp versions dictated by Genbank since these are constantly fluctuating due to consecutive attempts at re-annotation by TIGR. Table 7 lists all the similarities between Bglu421 and phiPasteur as well as the gi and genome map gp numbers. There were two other hypothetical conserved proteins similar between Bglu421 and phiPasteur, whose relationships were derived from high similarity degrees to an intermediate homolog in the database, connecting the two due to

the fact that the phiPasteur protein in question was one of the proteins newly added after re-annotation. Bglu421Dgp01and phiPasteurgpE were associated through similarity to *Xylella fastidiosa* 9a5c gi|9107484|, 4.00E-42 and 6.00E-73 respectively; Bglu421Dgp02 and phiPasteurgpD through shared resemblance to *Xylella fastidiosa* 9a5c gi|9107487| by expect values of 4.00E-55 and 1.00E-78. The blatant sequential organization of groups of these hypothetical proteins suggests that they are modules and were exchanged between both phages as such.

Three tail assembly proteins and three morons of Bglu421 bore similarities to phiPasteur proteins. The annotated baseplate assembly protein V (Bglu421Bgp27) had a degree of similarity of 5.00E-19 to *Acidiphilium cryptum* JF-5 gi|88939381| to which phiPasteurgpH had a resemblance of 2.00E-25, thus indicating that Bglu421gp27 is a homolog of phiPasteurgpH. The tail protein (Bglu421Bgp24) had direct similarity of 5.00E-37 to phiPasteurgp06. The phage tail sheath protein [17] of Bglu421 contig B resembled phiPasteurgp10 by an expect value of 1.00E-23.

The moron, Bglu421Cgp01 metal-dependent amidase/ aminoacylase/ carboxypeptidase was annotated as such by its strong similarity (5.00E-17) to phiPasteurgp35. Bglu421Bgp21, methionyl-tRNA synthetase resembles phiPasteurgp08 by a degree of 2.00E-20. Bglu421Bgp09 is a homolog to phiPasteurgp19 methyl-accepting chemotaxis protein by expect value of 3.00E-88. Here, the exchange of modules did not have a structural purpose, but instead seems to influence the phage's ability to survive within its bacterial host and/or the lysogen's ability to survive within its own host – consistent with moron characteristics.

Overall, it is clearly evident that the strongly clustering Bglu421protein groups to those in phiPasteur clearly indicate a modular exchange as well (see Fig. 8)– although not as strongly in the structural components of the phages as was seen in comparing Bcep176 to phiE125 and phi1026b in Figure 7.

Comparison of all phages. In comparing all five genomes (Bcep176, Bglu421, phiE125, phi1026b, and phiPasteur) there seemed to be a similar genome organization: tail assembly cassette, capsid assembly cassette, and lysis cassette. However, a discrete

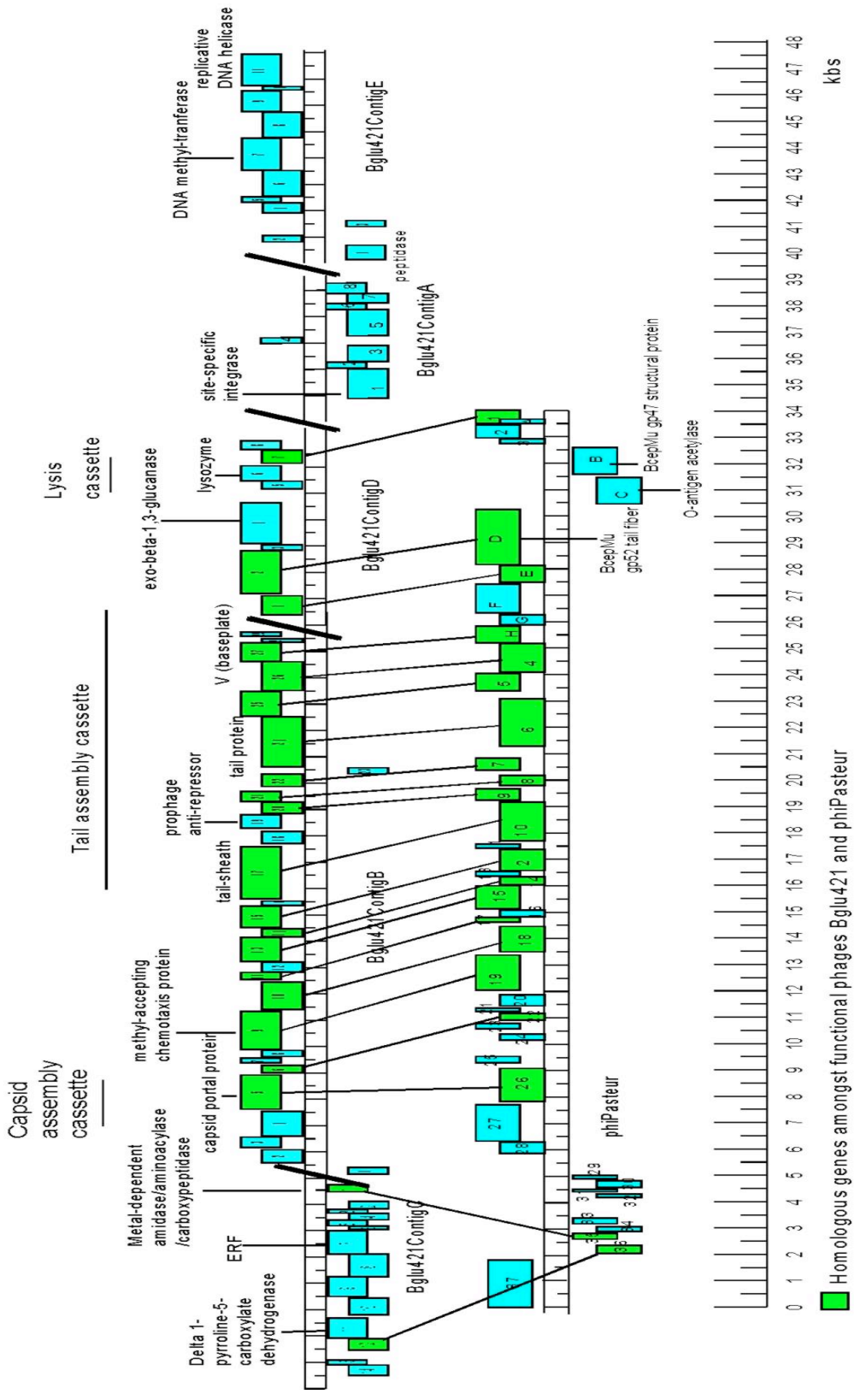

FIG. 8. Mosaicism (Bglu421, phiPasteur)

organization as such seen in phiE125 and phi1026b as well as Bcep176 is not clear in neither Bglu421 nor phiPasteur. (See Fig. 9)

There were similarities found between Bglu421 proteins and phi1026b and phiE125 proteins, although not in any way indicative of extensive modular exchange: site-specific integrase Bglu421Agp01 has a similarity of 2.00E-81 to phi1026bgp33 and 9.00E-80 to phiE125gp34; the hypothetical conserved protein Bglu421gp02 has a resemblance factor of 8.00E-9 to phi1026bgp34 and 3.00E-9 to phiE125gp35; hypothetical conserved protein Bglu421Agp05 has a marked similarity of 6.00E-105 to phi1026bgp38 and 3.00E-119 to phiE125gp37; peptidase Bglu42Egp01 is similar by an expect of 5.00E-7 to phiE125gp52 but has no counterpart in phi1026b; Bglu421Egp08, a PAPS reductase, also has no homolog in phi1026b but does have a strong resemblance of 2.00E-144 to phiE125gp57; lastly, Bglu421Egp09, a putative chromosome portioning protein SpoJ, has a similarity of 3.00E-65 to phi1026bgp68 and one of 1.00E-64 to phiE125gp58. Similarities across Bglu421 and phiE125 and/or phi1026b lay mostly in the morons region of both phages (see Table 8).

It is noteworthy, though disappointing, that Bglu421 had only one gene product with a counterpart Bcep176: hypothetical protein Bglu421Cgp05 to Bcep176gp38 with an expect value of 8.00E-7.

phiPasteur, on the other hand, had two homologs to proteins in Bcep176: SAR endolysin phiPasteurgp02, closely similar by 9.00E-54 to homolog Bcep176gp58 (homolog of phiE125gp25 and phi1026bgp24), and hypothetical conserved protein phiPasteurgp33 similar by 1.00E-14 to homolog Bcep176gp39 – neither of these connections are depicted in Figure 9.

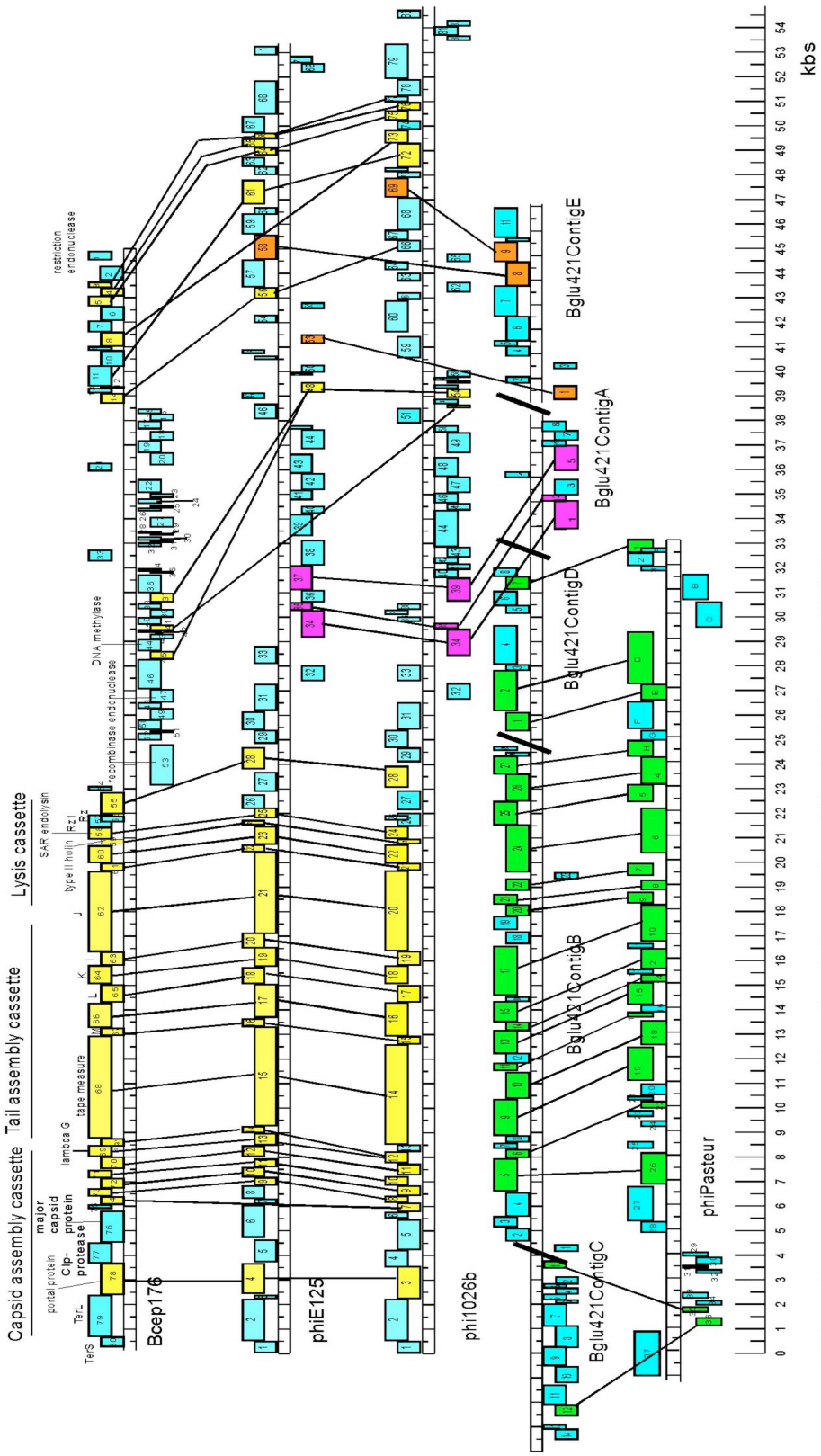

FIG. 9. Overall Mosaicism (Bcep176, phiE125, phi1026b, Bglu421, phiPasteur)

TABLE 8. Bglu421 proteins to phi1026b and phiE125 mapped proteins

Protein in Bglu421	map name in phi1026b/gi(gp)	map name in phiE125/gi(gp)	protein encoded
Bglu421Agp01	p34/gi|38707923(gp33)	p34/gi|17975195(gp34)	site-specific integrase
Bglu421Agp02	p35/gi|38707924(gp34)	p35/gi|17975196(gp35)	hypothetical conserved protein
Bglu421Agp05	p39/gi|38707928(gp38)	p37/gi|17975198(gp37)	hypothetical conserved protein
Bglu421Egp01	none	p53/gi|17975214(gp52)	peptidase
Bglu421Egp08	none	p58/gi|17975219(gp57)	PAPS reductase
Bglu421Egp09	p69/gi|38707958(gp68)	p59/gi|17975220(gp58)	putative chromosome partitioning protein SpoJ

CONCLUSIONS

Just one moron from a lysogenic phage can, through lysogenic conversion, cause its bacterial host to become pathogenic or increase its fitness as a pathogen (3).

In analyzing Bcep176, gene products of phages phiE125 and phi1026b have an obvious mosaic relationship to those found in Bcep176. phiE125 is phage isolated from *Burkholderia thailandensis*, a non-pathogenic bacteria (26) well-known for its close relatives *B. mallei* and *B. pseudomallei* (possibly misused as biological weapons (7,8). Previous work on phiE125 identified it as a siphophage and revealed that it possessed lambdoid morphology in its capsid and tail assembly cassettes (26). phi1026b, on the other hand, was isolated from *B. pseudomallei* 1026b in an effort to use the phage as a diagnostic tool to differentiate between the two closely related biological threat agents *B. pseudomallei* and *B. mallei* (13). In this same experiment, phi1026b and was shown to possess near-exact mosaicism with phiE125 in its tail, packaging, host lysis, and DNA replication gene products (13).

The host of Bcep176, *B. multivorans*, is believed to be involved in pulmonary infections (9). The strong connections between Bcep176 and phiE125 and phi1026b particularly in head-to-tail proteins and tail assembly cassette indicates that Bcep176 is, by association, a lambdoid phage as well. Bcep176's connection to cystic fibrosis (9) and *B. pseudomallei* and *B. mallei*'s use as biological weapons, bring to the light the possibility that perhaps a specific subgroup of tail proteins in lambdoid phage make the phages more fit to infecting *B. multivorans*. If this is true, further studies connecting the three phages can bring us closer to finding connections between *B. multivorans* and its other cystic fibrosis-causing sister strains (23) and *B. cepacia* cousins (11), and perhaps even help distinguish what helps *Burkholderia* bacteria make the leap from soil to airborne.

The evident modular mosaicism between phiPasteur and Bglu421 brings to question the connection between these two phages of pathogens from entirely different ends of the spectrum: *B. pseudomallei*, as previously mentioned, is an animal pathogen

that causes melioidosis in animals and humans (12) while Bglu421 is a plant pathogen that causes seedling rot (blight) in rice (19). Interestingly enough, although *B. pseudomallei* can be isolated from water and soil surfaces in endemic areas, it is also easily isolated from rice paddies (21). Since the strain of *B. glumae* used in the laboratory was isolated from rice (Gonzalez, TX) then the mosaicism evident between the two may have been propitiated by the plausible proximity of possible reservoirs of both bacteria: rice and rice paddies.

One particular region of similarity in Contig B of Bcep176 in relation to phiPasteur revealed an interesting difference between the two: phiPasteur has a possible animal pathogen pathogenicity factor, the o-antigen acetylase [C], but in its corresponding place on the Bglu421 genome there is a predicted exo-beta-1, 3-glucanase at Bglu421Dgp04. This particular difference between the two indicates that the glucanase protein is in a location permissible to exchanging morons. The assumption that phiPasteurgpC is a pathogenicity factor can only further suggest that Bglu421Dgp04 could be a moron that contributes a pathogenicity factor involved in *Burkholderia glumae*'s fitness as a rice pathogen. This preliminary analysis of Bglu421 is only a sample of the type of information that could be gleaned from further studying the bacteria and the phage to elucidate on the phage-bacteria-rice interaction, and lead to improvements in rice blight management and prevention.

Further testing of phiPasteur would be required in order to determine if it is functional and whether or not the O-antigen acetylase [C] does contribute to *B. pseudomallei* phiPasteur's pathogenicity. Completion and ample coverage of Bglu421 genome is essential in order to elucidate more conclusive evidence of its suspected siphophage morphology and experiments need to be conduced in order to determine the biochemical pathways and function of its many enzymes, in particular that of the exo-beta-1, 3-glucanase of Contig D.

REFERENCES

1. **Barrow, P. A. and J. S. Soothill**. 1997. Bacteriophage therapy and prophylaxis: rediscovery and renewed assessment of potential. Trends in Microbiol. **5**(7):268-271.
2. **Botstein, D.** 1980. A theory of modular evolution for bacteriophages. Ann. NY Acad. Sci. **354**:484-490.
3. **Brüssow, H., C. Canchaya, W. Hardt**. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. Micr. Mol. Biol. Reviews **68**(3):560-602.
4. **Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW**. 2005. The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. J. Bacteriol. **187**(3):1091-104.
5. **Casjens, S.** 2003. Prophages and bacterial genomics: what have we learned so far?. Mol. Microbiol. **49**(2):277-300.
6. **Catalano, C. E., D. Cue, and M. Feiss.** 1995. Virus DNA packaging: the strategy used by phage lambda. Mol. Microbiol. **16:1075-1086**
7. **Centers for disease Control and Prevention (CDC)**. 2000. Biological and chemical terrorism: strategic plan for preparedness and response. Recommendations of the CDC Strategic Planning Workgroup. MMWR Morb. Mortal. Wkly. Recomm. Rep. Apr 21; **49**(RR-4):1-14.
8. **Centers for disease Control and Prevention (CDC).** 2004. Laboratory exposure to *Burkholderia pseudomallei*--Los Angeles, California, 2003. MMWR Morb. Mortal. Wkly. Rep. Oct 29;**53**(42):988-90.
9. **Chu, K. K., K. L. MacDonald, D. J. Davidson, and D. P. Spencert.** 2004. Persistence of *Burkholderia multivorans* within the Pulmonary Macrophage in the Murine Lung. Infect. Immun. **72**: 6142–6147.

10. **Coenye, T., P. Vandamme, J. R. W. Govan, and J. J. LiPuma**. 2001.Taxonomy and Identification of the *Burkholderia capeacia Complex*. J. Clin. Microbiol. **39**: 3427-3436.
11. **Cunha, M. V., J. H. Leitao, E. Mahenthiralingam, P. Vandamme, L. Lito, C. Barreto, M. J. Salgado, and I. Sa-Correia**. 2003. Molecular analysis of *Burkholderia cepacia* complex isolates from a Portuguese cystic fibrosis center: a 7-year study. J. Clin. Microbiol. **41**:4113-4120.
12. **Dance, D. A. B**. 2006. Melioidosis, p.381-388. *In* R. L. Guerrant, D.H. Walker, P. F. Weller (ed.). Tropical Infectious Diseases: principles, pathogens and practice, 2nd ed., vol.1. Elsevier Churchill Livingstone, Philadelphia.
13. **DeShazer, D**. (2004). Genomic diversity of *Burkholderia pseudomallei* clinical isolates: subtractive hybridization reveals a *Burkholderia mallei*-specific prophage in B. *pseudomallei* 1026b. J. Bacteriol. **186**:3938-50.
14. **Godoy, D., G. Randle, A. J. Simpson, D. M. Aanensen, T. L. Pitt, R. Kinoshita, and B. G. Spratt**. 2003. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. J. Clin. Microbiol. **41**:2068-2079.
15. **Goris, J., P. De Vos, J. Caballero-Mellado, J. Park, E. Falsen, J. F. Quensen III, J. M. Tiedje, and P. Vandamme.** 2004. Classification of the PCB- and biphenyl degrading strain LB400 and relatives as *Burkholderia xenovorans* sp. nov. Int. J. Syst. Evol. Microbiol. **54**:1677-1681.
16. **Groman, N.B.** 1984. Conversion by corynephages and its role in the natural history of diphtheria. J. Hyg. (Lond). **93**(3):405-17.
17. **Hershey, A.D.** 1971. The bacteriophage lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
18. **Hendrix, R. W.** 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. **61:**471-480.
19. **Li, H. Q., I. Matsuda, Y. Fujise, and A. Ichiyama**. 1999. Short-chain acyl-CoA-dependent production of oxalate from oxaloacetate by *Burkholderia glumae*, a plant

pathogen which causes grain rot and seedling rot of rice via the oxalate production. J. Biochem. Tokyo 126:243-53.

20. **Plunket III, G., D. J. Rose, T.J. Durfee, and F. R. Blattner**. 1999. Sequence of Shiga Toxin 2 Phage 933W from *Escherichia coli* O157:H7: Shiga Toxin as a Phage Late-Gene Product. J. Bacteriol. **181**(6): 1767-78.
21. **Strauss, J. M., M. G. Groves, M. Mariappan, D. W. Ellison**. 1969. Melioidosis in Malaysia. II. Distribution of Pseudomonas *pseudomallei* in soil and surface water. Am. J. Trop. Med. Hyg**. 18**(5):698-702.
22. **Susskind, M**. **M. and D. Botstein.** 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. **42:**385-413.
23. **Vandamme, P., B. Holmes, M. Vancanneyt, T. Coenye, B. Hoste, R. Coopman, H. Revets, S. Lauwers, M. Gillis, K. Kersters, and J. R. W. Govan.** 1997. Occurrence of Multiple Gemnovars in Cystic Fibrosis Patients and Proposal of *Burkholderia multivorans* sp. nov. Int. J. of Sys. Bacter. **47**(4)**:**1188-1200.
24. **Waldor, M. K. and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272**(5): 1910-1914.
25. **Westmoreland, B.C., W. Szybalski, and H. Ris.** 1969. Mapping of deletions and substitutions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy. Science **163:**1343-1348.
26. **Woods, D. E., J. A. Jeddeloh, D. L. Fritz, and D. DeShazer**. (2002)*Burkholderia thailandensis* E125 harbors a temperate Bacteriophage specific for *Burkholderia mallei*. J. Bacteriol. **184**:4003-4017.

Supplemental sources consulted

- Büchen-Osmond, C. 2004. 02.066.0.01.001. accessed January 25, 2006. Enterobacteria phage λ. *In:* ICTVdB - The Universal Virus Database, version 3. ICTVdB Management, Columbia University, NY. http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/

CONTACT INFORMATION

Name: Linet Mera

Address: Department of Biochemistry and Biophysics
Texas A&M University
College Station, TX 77845

E-mail address: linet.mera @ gmail.com

Education: B.S. Genetics and Biochemistry; Texas A&M University at College Station, 2006